Genetic variants within glutathione *S*-transferase alpha 4 and their role in newly-diagnosed epilepsy

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Master of Philosophy

Paul Stephen Mitchell

August 2012

Declaration

This thesis is the result of my own work. The material contained within this thesis has not been presented, nor is it currently being presented, either wholly or in part for any degree or other qualification.

Paul Stephen Mitchell

This research was conducted at the Wolfson Centre for Personalised Medicine

Department of Molecular and Clinical Pharmacology

University of Liverpool

Table of Contents

Contents

List of Tables

Table of Figures

Acknowledgements

Foremost, I would like to offer my sincerest gratitude to my supervisors, Dr Graeme Sills and Professor Tony Marson. I am indebted to their patience and academic guidance during my year of postgraduate research, and for this I am wholeheartedly grateful – I am confident the skills I have gained this year will serve me well for my future medical career.

I would also like to extend great thanks to Dr Dan Carr for his infinite patience and support in every aspect of this project; for never once judging any outrageous nonsense I came out with, and for teaching me the genotyping methodologies required for this MPhil.

For input into the statistical analyses, I would like to thank Dr Andrea Jorgensen, whose help and guidance in making sense of medical statistics has been wholly appreciated.

To everyone else in the departments of neuroscience and molecular and clinical pharmacology, I extend my thanks for their contributions.

Finally, I would like to thank my friends and family for their continued support in my academic career. They have helped guide every decision throughout my medical and masters studies, and without whom I would undoubtedly not be where I am today.

Abbreviations

µL	Microlitre
µM	Micromole
3’/5’ UTR	3’/5’ untranslated region
4-HNE	4-hydroxynonenal
A	Adenine
AED(s)	Antiepileptic drug(s)
ALT	Alanine aminotransferase
Ask1	Apoptosis signal-regulating kinase 1
AST	Aspartate aminotransferase
AUS	Australia
C	Cytosine
cDNA	Complementary deoxyribonucleic acid
CI	Confidence interval
CT	Computed tomography
CYP2C9	Cytochrome P450 2C9
DNA	Deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DVLA	Drive and vehicle licencing agency
E.coli	Escherichia coli
EEG	Electroencephalogram
EST	Expanded sequence tag
FAM	6-carboxyflurorescein
FDR	False discover rate
FLAP	5-lipooxygenase activating protein
G	Guanine
GBP	Gabapentin
GSH	Glutathione
GSSH	Glutathione disulphide
GST	Glutathione S-transferase

GSTA	Glutathione S-transferase alpha
GSTM	Glutathione S-transferase mu
GSTO	Glutathione S-transferase omega
GSTP	Glutathione S-transferase pi
GSTS	Glutathione S-transferase sigma
GSTT	Glutathione S-transferase theta
GSTZ	Glutathione S-transferase zeta
GWAS	Genome wide association study
GWS	Genome wide significance
HNF-1	Hepatic nuclear factor 1
HR	Hazard ratio
HW	Hardy-Weinberg equilibrium
IGE	Idiopathic generalised epilepsy
IL-1β	Interleukin 1β
ILAE	International league against epilepsy
JNK	c-Jun N-terminal kinase
Kb	kilobase
KM	Kaplan-Meir
kNN	k nearest neighbour
LD	Linkage disequilibrium
LRE	Localisation related epilepsy
LTC_4S	Leukotriene C_4 synthase
MAAI	Maleylacetoacetate isomerase
MAF	Minor allele frequency
MALDI-TOF MS	Matrix assisted laser desorption/ionisation-time of flight mass spectrometry
MAPEG	Membrane associated proteins in eicosanoid and glutathione metabolism
MDA	Malondialdehyde
mEH (3/4)	Microsomal epoxide hydrolase (3/4)
$MgCl_2$	Magnesium chloride
MGST	Microsomal glutathione S-transferase

MGST(1/2/3)	Microsomal glutathione S-transferase (1/2/3)
MLH	Matrix liquid handler
MRI	Magnetic resonance imaging
mRNA UTR	Messenger ribonucleic acid untranslated region
mRNA	Messenger ribonucleic acid
NAPQI	*N*-acetyl-*p*-benzoquinone imine
NCBI	National centre for biotechnology reseach
ng	Nanograms
NHS	National health service
NICE	National institute of clinical excellence
^{O}C	Degrees centigrade
OR	Odds ratio
PAH	Polyaromatic hydrocarbon
PCR	Polymerase chain reaction
PGD_2	Prostaglandin D_2
PGES	Prostaglandin E synthase
PGF_2	Prostaglandin F_2
PGH_2	Prostaglandin H_2
q-PCR	Quantitative polymerase chain reaction
rpm	Revolutions per minute
SANAD	Standard and new antiepileptic drug
SAP	Shrimp alkaline phosphatase
SD	Standard deviation
SNP	Single nucleotide polymorphism
SUDEP	Sudden unexplained death in epilepsy
T	Thymine
UNC	Unclear
UV	Ultraviolet
VKORC1	Vitamin K epoxide reductase complex subunit 1
VPA	Valproic acid/Sodium valproate
γ-GT/GGT	Gamma glutamyl transpeptidase

Abstract

Introduction: Epilepsy is one of the commonest neurological disorders, affecting over 50 million people worldwide. Antiepileptic drugs (AEDs) are the mainstay of treatment, though they remain ineffective in 30% of individuals for whom they are prescribed.

Pharmacoresistance in epilepsy is a significant problem. The mechanisms involved elude us still, with numerous theories and hypotheses posed in order to explain why this phenotype may exist. It is thought that genetic variation within certain genes may explain, at least in part, why the pharmacoresistant phenotype exists, and indeed may explain the dose-response variation observed in clinical practise.

A genome-wide association study was conducted on newly-diagnosed patients with epilepsy upon starting their first AED, with the primary measure of outcome being one year remission of seizures at any time during follow-up. The strongest association was seen at a loci on chromosome 6, rs622902 ($p=1.47\times10^{-7}$) in the glutathione *S*-transferase alpha 4 (*GSTA4)* gene. GSTA4 is an enzyme of the glutathione-conjugating family of superenzymes that are responsible for a plethora of reactions that detoxify reactive metabolic intermediates, including many xenobiotics. In view of these findings, it was the aim of this study to explore the role of GSTA4 in newly-diagnosed epilepsy patients.

Methods: Seventeen tagging SNPs representing 48 alleles across the *GSTA4* gene were genotyped in 541 SANAD and 390 Glasgow patients using TaqMan quantitative PCR and Sequenom matrix-assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-TOF MS). Binary logistic regression and Cox proportional hazards multiple regression analyses were conducted in order to analyse patient remission (yes/no) of epilepsy, and time to 12-month remission of epilepsy, respectively. Six clinical covariates were adjusted for in the analysis: age, sex, epilepsy type (partial, generalised, unclassifiable), number of pre-treatment seizures (2, 3, 4, 5, >5), AED

treatment (sodium valproate or gabapentin vs. neither) and EEG findings (normal, non-specific, epileptiform, not-done).

Results: Three SNPs (rs6904769, FDR = 0.0353; rs6922246, FDR = 0.0109 and rs384505, FDR = 0.0459) in the SANAD cohort showed an association with the binary outcome (yes/no) of seizure remission. Only rs6922246 (FDR = 0.0459) held association after combining both cohorts.

Two SNPs (rs6904769, FDR = 0.0491 and rs6922246, FDR = 0.00380) were significantly associated with time to 12-month remission of epilepsy. This association held in rs6922246 (FDR = 0.0459) after combining the cohorts.

Conclusion: Polymorphism in rs6922246 may be associated with both the remission of epilepsy and the time to 12-month remission of epilepsy. Validation in an independent cohort is necessary to explore this link further.

Chapter 1

Introduction

Chapter 1: Introduction

1.1 Epilepsy

As with many diseases - particularly those affecting the nervous system - epilepsy was considered a symptom of demonic possession. For epilepsy, this belief was held firmly until the period 1860-1910, when the supernatural explanation for the cause of epilepsy was lain to rest, and the modern era of scientific theory was initiated by the 'father of epilepsy', John Hughlings Jackson (1).

Much like headache, epilepsy is a symptom of neurological dysfunction rather than a disease in itself per se (2). The distinction between a seizure and epilepsy is subtle, in that any person can have a seizure under certain pathological conditions such as hyponatraemia, hypoglycaemia, alcohol withdrawal, etc., whereas epilepsy is a condition of recurrent, unprovoked seizures whereby the known pathological conditions such as those aforementioned are absent. Therefore, every patient with epilepsy has seizures, but not ever patient with seizures has epilepsy. The underlying neurological pathology in epilepsy is manifest as spontaneous neuronal discharge of cerebral cortex neurons resulting in seizures. The epilepsy type is dependent on a variety of factors, though the anatomical location of the epileptogenic focus, along with any secondary generalisation, dictates the phenotype most strongly.

The epilepsies comprise a group of disorders which are diagnosed following two or more spontaneous seizures for which no other precipitating factor may be attributable – that is, having ruled out cardiac, metabolic and other systemic causes for seizures. Thorough neurological assessment to include neuroimaging is necessitated in certain patient groups depending on the list of likely differential diagnoses, as seizures may be secondary to the presence of other central nervous system pathology (space occupying lesions), or may ensue following a 'normal' physiological reaction to fever – i.e., childhood febrile seizure. Care must be taken with the diagnosis, as its label is often commensurate with

negative connotations: quality of life is often greatly impacted, and stigma is something many patients with epilepsy often report.

Quality of life in epilepsy is negatively impacted upon for several reasons. Unlike other chronic diseases, epilepsy may result in withdrawal of driving privileges, such that one must be seizure-free for 12 consecutive months in order to regain a licence from the Drive and Vehicle Licensing Agency (DVLA). A measurement of outcome in epilepsy research is time to 12-month remission, which ideally encompasses this restriction, also. Furthermore, the inability to predict the next seizure often leads to patients being cautious about being away from home, developing anxiety and becoming caught in a spiral of fear and emotion often resulting in depression as a concurrent diagnosis.

A multitude of underlying factors are recognised as responsible for the causation of epilepsy, and it is now thought that the disease is multifactorial in most patients. The list of aetiologies is vast, and can range from underlying genetic factors through to structural cerebral pathology as in mesial temporal sclerosis. Sclerosis of the mesial temporal lobe structures is not considered to be primary causative, as it is not strictly known whether the sclerosis arises as a result of the epilepsy, or whether the sclerosis is a prerequisite to the epilepsy. This may be the case for other pathologies observed in patients with epilepsy.

The mainstay of medical treatment is with antiepileptic drugs (AEDs). These drugs are successful in treating the epilepsy in around 70% of patients who take them, with the remaining 30% being considered pharmacoresistant (3, 4). Pharmacoresistance in epilepsy yields yet more problems for patients, as poorly-controlled seizures may result in significant morbidity during a loss of consciousness, or even mortality as in the sudden unexplained death in epilepsy (SUDEP).

1.1.1. Incidence and Prevalence

Epilepsy is a common neurological condition affecting over 50 million people worldwide, and this figure is a likely gross underestimate (5). Unlike many chronic diseases, epilepsy is, for the most part, a hidden disorder, and is more widespread in society than is often thought. The incidence of epilepsy as cited in the literature varies significantly, though a large meta-analysis in 2011 by Ngugi et al (5) reported that the mean incidence was 50.4 cases per 100,000 population per annum. This figure is higher for those living in developing nations (81.7 per 100,000 per year) and is also higher for those of lower socioeconomic status. The prevalence of the disorder is estimated at between 4 and 10 persons per 1,000 (6). Increased prevalence of epilepsy in developing nations is most likely the result of poor maternal health, particularly in the first trimester during neural tube development, higher birth complications and subsequent neonatal anoxia, and poorer environmental conditions during the perinatal period (5, 6).

The distribution of incidence against age for epilepsy is bimodal; the first peak is in the early years of life, and the second occurs in the elderly population. The latter probably occurs in accordance with a greater incidence of neurological dysfunction with advancing age, most notably cerebrovascular disease. In recent years, the incidence of epilepsy in the paediatric population has seen a decline, and this is thought to have occurred secondary to better maternal heath during pregnancy and increased standards of neonatal and perinatal care.

1.1.2. Aetiology of epilepsy

Unfortunately, the aetiologies of epilepsy, on the whole, are still poorly understood. As with many conditions, notions regarding causation have evolved over time in tandem with our scientific knowledge about the physiology of health and disease. This is particularly true for epilepsy. The advent of advanced neuroimaging, both computed tomography

(CT) and magnetic resonance imaging (MRI), along with significant advances in molecular biochemistry, have revolutionised epileptology, though there is still a long way to go. It is now possible to discern structural cortical abnormality from other causes of epilepsy, and many of the inherited errors of metabolism that lead to specific epilepsy syndromes are well understood (1). However, the vast majority of epilepsies still remain 'cryptogenic.' Genetic predisposition is thought to contribute significantly, and recent advances in molecular genetics offer further light and promise to understanding the aetiology of epilepsy.

1.1.3. Diagnosis

The diagnosis of epilepsy is notoriously difficult. As such, it is a recommendation of the National Institute for Health and Clinical Excellence (NICE) that the diagnosis be confirmed by a specialist clinician with an interest in the condition (7).

Whilst the majority of medical diagnoses can be made on history and examination, clinical investigations such as haematological, biochemical and/or imaging studies can often confirm or refute the diagnosis. This is often not the case with epilepsy. A thorough history of the seizure/s, including details of the events preceding and following the seizure, is vitally important. A collateral history from an eye-witness is also an invaluable part of the diagnostic process. Neuroimaging or neurophysiological testing may offer little additional information to the initial event, and clinical examination is often unremarkable. Once other causes have been ruled out, a diagnosis of epilepsy can be made following two or more unprovoked seizures, and only then do other studies such as electroencephalography (EEG) or neuroimaging become useful for confirmatory or classification purposes.

1.1.4. The classification of seizures and epilepsy

Seizures are clinically classified into two main categories: partial/localised/focal and generalised. Where the seizure cannot be classified, it should be labelled 'unclassifiable' until such time as additional information becomes available enabling its inclusion in one of the two principal categories (8).

This classification system is centred on the anatomical area of cortical discharge that manifests as a seizure. Partial seizures are those that occur with a discrete area of discharge that is attributable to a particular area of brain and that does not involve the entire cerebral hemispheres bilaterally. Generalised seizures involve all areas of the brain having spread via the deep brain structures/corpus callosum to involve both hemispheres. The focus in partial seizures is often easily discerned by the seizure itself; involvement of the motor strip will affect the contralateral limb(s), and activation of the reticular formation in the midbrain will lead to loss of consciousness, as in generalised seizures or complex partial seizures.

Within these two broad categories, there are several subtypes that are based upon their clinical manifestations. The International League Against Epilepsy (ILAE) published two systems of classification: one regarding the classification of seizures, and the other on epilepsies. The seizure classification is outlined in a simplified manner in table 1.1 below, and the epilepsy classification is detailed thereafter. A detailed description of epilepsy subtypes and syndromes is beyond the scope of this short introduction.

Table 1.1. Simplified adaption from the 1981 ILAE classification of seizures(8).

Classification of Seizure	Seizure Type
Generalised (consciousness impaired)	Tonic-clonic Absence Myoclonic Tonic Clonic Atonic
Partial	Simple partial Complex partial (consciousness impaired) Partial with secondary generalisation (consciousness impaired)
'Unclassifiable'	

Below is the latest accepted classification of the epilepsies from the ILAE.

1.Localization-related (local, focal, partial) epilepsies and syndromes

1.1 Idiopathic (with age-related onset)

Benign childhood epilepsy with centro-temporal spikes

Childhood epilepsy with occipital paroxysms

Primary reading epilepsy

1.2 Symptomatic

Chronic progressive epilepsia partialis continua

Syndromes characterized by seizures with specific modes of precipitation

Temporal lobe epilepsies

Frontal lobe epilepsies

Parietal lobe epilepsies

Occipital lobe epilepsies

1.3 Cryptogenic

2. Generalized epilepsies and syndromes

2.1 Idiopathic (with age-related onset)

Benign neonatal familial convulsions

Benign neonatal convulsions

Benign myoclonic epilepsy in infancy

Childhood absence epilepsy

Juvenile absence epilepsy

Juvenile myoclonic epilepsy (impulsive petit mal, Janz syndrome)

Epilepsy with grand mal seizures (GTCS) on awakening

Other generalized idiopathic epilepsies

Epilepsies with seizures precipitated by specific modes of activation

2.2 Cryptogenic or symptomatic

West syndrome (infantile spasms)

Lennox-Gastaut syndrome

Epilepsy with myoclonic-astatic seizures

Epilepsy with myoclonic absences

2.3 Symptomatic

2.3.1 Non-specific aetiology

Early myoclonic encephalopathy

Early infantile epileptic encephalopathy with suppression bursts

Other symptomatic generalized epilepsies

2.3.2 Specific syndromes

Epileptic seizures complicating other disease states

1.1.5. Pharmacological treatment of epilepsy

Pharmacological therapy remains the mainstay of epilepsy treatment. The ultimate goal of AED therapy is to achieve total seizure freedom whilst producing no side-effects from the medication. Seizure freedom with no side-effects is a scenario seldom seen by the clinician, and often a balance between seizure control and side-effects must be struck in order to obtain the best possible outcome for any given patient.

Of patients treated with AEDs, approximately 70% will have their seizures well-controlled, while the remaining 30% of patients are considered to have pharmacoresistant epilepsy and receive little (if any) benefit from drug therapy (3, 4).

The current market has over 20 AEDs available for prescription. Figure1.1 illustrates the main AEDs available today and their approximate year of approval for marketing. Even with the exponential increase in AEDs coming to market in recent decades, there has been no real revolution in the treatment of the condition (9). With approximately 50 million people worldwide suffering from epilepsy and 30% of those inadequately served by existing medication, pharmacoresistant epilepsy represents a significant global health issue.

There is a clear need for the development of novel, more effective AEDs. However, this is unlikely to be achieved with any significant success until the mechanisms underpinning epileptogenesis are unravelled and new targets for drugs are discovered.

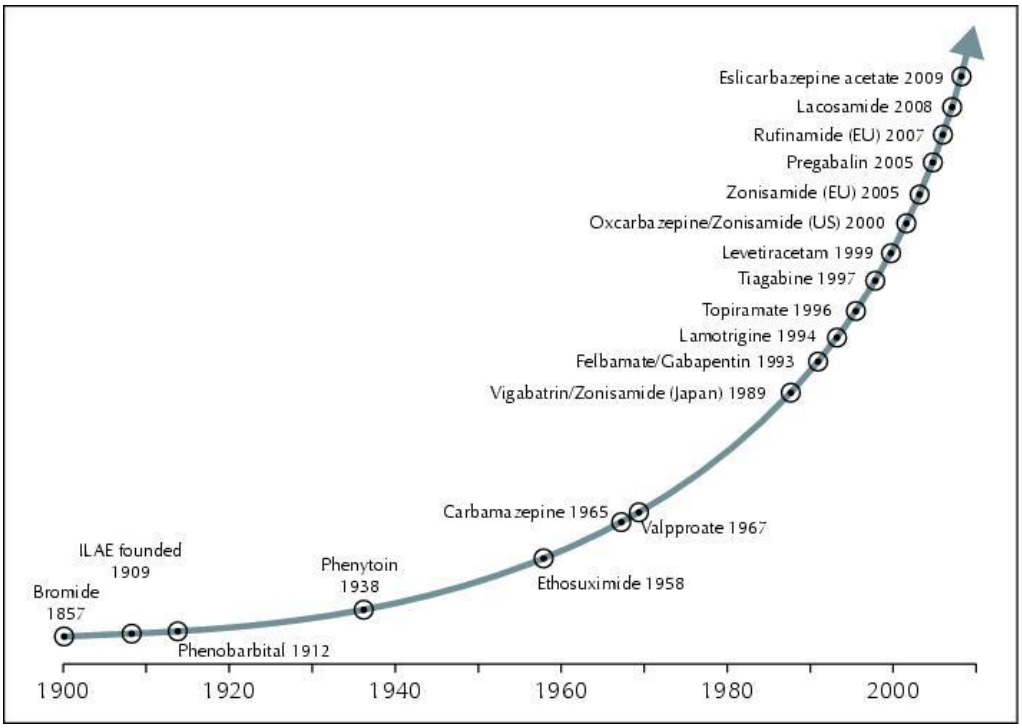

Figure 1.1. AED development since 1900, showing the exponential rise in AED drugs coming to market in the past 20 years.
Taken from Arzimanoglou et al (9).

1.1.5.1. Choice of antiepileptic medication

The choice of AED is a decision directed by guidelines published by NICE (7), which reflect evidence-based practise and the opinion and experience of specialist clinicians with an interest in epilepsy. Several factors are considered in AED prescribing; foremost is the suitability of the drug for a specific seizure type or syndrome. Most AEDs are not suitable for all seizure/epilepsy types, and in some circumstances may even exacerbate the condition. The side-effect profile of the medication must also be borne in mind, with switching of AEDs commonplace owing to unacceptable side-effects. Further factors influencing choice include age, sex, concomitant medications, patient preference (usually on the basis of anticipated side effects), and cost.

The current guidelines for AED selection published by NICE are summarised in table 1.2 below. Specific epilepsy syndromes and status epilepticus are not included here. It is advised that monotherapy be used wherever possible, and that add-on therapy should not be attempted until all first-line agents have been trialled to their maximum tolerated dose.

Table 1.2: Utilisation of AEDs as recommended by NICE. Adapted from NICE (7).

Seizure Type	1st-line AED	Adjunctive AED	Others
Generalised tonic-clonic	Carbamazepine Lamotrigine Oxcarbazepine Sodium valproate	Clobazam Lamotrigine Levetiracetam Sodium valproate Topiramate	
Tonic or atonic	Sodium valproate	Lamotrigine	Rufinamide Topiramate
Absence	Ethosuxamide Lamotrigine Sodium valproate	Ethosuxamide Lamotrigine Sodium valproate	Clobazam Clonazepam Levetiracetam Topiramate Zonisamide
Myoclonic	Levetiracetam Sodium valproate Topiramate	Levetiracetam Sodium valproate Topiramate	Clobazam Clonazepam Piracetam Zonisamide
Focal	Carbamazepine Lamotrigine Levetiracetam Oxcarbazepine Sodium valproate	Carbamazepine Clobazam Gabapentin Lamotrigine Levetiracetam Oxcarbazepine Sodium valproate Topiramate	Eslicarbazepine acetate Lacosamide Phenobarbital Phenytoin Pregabalin Tiagabine Vigabatrin Zonisamide

Chapter 2

Early work, patient cohorts, and genome wide association study

Chapter 2: Early work, patient cohorts, and genome wide association study

2.1. The SANAD cohort

The standard and new antiepileptic drug (SANAD) study was an un-blinded, multicentre, randomised clinical trial designed to analyse which of the current standard and new antiepileptic drugs had superiority in both clinical and cost effectiveness in newly-diagnosed patients with epilepsy. Over 2,400 patients were recruited into the trial, and were prospectively followed up for a period of at least 1 year from the initiation of their first AED (10, 11). Drug response phenotypes were reported by consulting the clinical trial database.

A subset of the SANAD cohort was consented to provide a blood sample that was to be used for DNA extraction and later genetic analysis. The approval for this was given by the North-West multicentre research ethics committee in August 2002.

2.2 The Glasgow cohort

The Glasgow cohort comprises 285 newly-diagnosed patients with epilepsy that were recruited from the Western Infirmary, Glasgow. Drug response phenotypes in this study were identified retrospectively from prospectively-collected trial data and/or from hospital notes, thereby making the available clinical data more comprehensive than the SANAD data set. Patients were self-reported as having European ancestry and blood samples for DNA analysis were collected with consent following approval from the West ethics committee, North Glasgow University Hospitals NHS Trust in September 2002.

2.3 The Melbourne cohort

The Melbourne epilepsy pharmacogenetics cohort (hereafter designated the 'AUS' cohort) was a prospective study of the genetic influences on treatment outcomes in newly-diagnosed epilepsy patients. This cohort consisted of 115 patients, in which 4,041

SNPs in 279 candidate genes were used in order to try and determine which SNPs were most influential in predicting seizure control following 1 year of AED therapy using a k-Nearest neighbour (kNN) model.

2.4. Variation between the SANAD and Glasgow cohorts

Differences between the cohorts that were of note include the geographical distribution of the patients in the SANAD cohort; these patients are representative of the UK with their diagnoses being made by clinicians nation-wide. Conversely, the Glasgow cohort was representative of a discrete geographical area with the diagnoses made by the same group of clinicians. Further, whilst the Glasgow data was deemed to be more comprehensive, data that was not available from this group included information regarding neurological deficit. The significance of these is discussed in the appropriate sections.

2.5. Genome wide association study

Genome wide association studies (GWASs) have revolutionised the experimental basis of how we study the genetics of human disease. This approach relies on data obtained by the International HapMap project, which is a multi-country effort to catalogue genetic similarities and differences within humans in order to better understand how genetics contributes to disease and variation in treatment responses with pharmaceutical agents. GWAS also takes advantage of the fact that variance within the genome at one particular locus can be used to predict variance at an adjacent locus. These probabilities remain high over distances of approximately 30,000 base pairs of DNA within the human genome, which enables the entire genome to be analysed for common variants by genotyping approximately 500,000 markers of the estimated 3-10 million polymorphisms believed to be present (12).

The results of a GWAS are typically illustrated in the form of a Manhattan plot, which displays statistical significance on the y-axis (as $-\log_{10}$ p-value) against chromosome number on the x-axis, for each genotyped SNP. Genome wide significance (GWS) is arbitrarily assumed at a p-value of $<1x10^{-4.5}$ for discovery analysis and a p-value of $<5x10^{-8}$for meta-analysis used in validation. Below GWS, the general consensus is that meta p-values in the interval $5x10^{-8} < p < 5x10^{-7}$ are suggestive of causal genomic association.

A GWAS of treatment response in newly-diagnosed epilepsy patients was conducted on patients from the SANAD cohort at the Wellcome Trust Sanger Institute on an Illumina 660Q GWAS chip (unpublished data). The aim of this was to identify regions of DNA that were highly polymorphic in order to begin to understand the role of SNP variations in the response to AED therapy. These data were meta-analysed with those from the AUS cohort for the purposes of validation. The primary outcome in this analysis was 1-year remission of seizures at any stage during follow-up, with a minimum follow-up of one year from initiation of AED therapy.

The analysis was adjusted for clinical covariates (listed below) that were discovered to have a significant ($p<0.05$) influence on the trait of interest (i.e., likelihood of 1-year remission) following a univariate logistic regression of the SANAD data set. Validation analysis in the AUS cohort was also adjusted for these covariates, regardless of whether they were significantly associated with outcome in this cohort alone.

2.5.1 Clinical covariates included in the GWAS analysis

The following is a list of the clinical covariates included in the GWAS analysis.

- Age
- Sex
- Epilepsy type (generalised, partial or unclassified)
- Treatment (sodium valproate, gabapentin or neither)
- EEG (normal, abnormal or not done)
- Neurological deficit (yes or no)
- Number of pre-treatment seizures (2, 3, 4, 5 or >5)

A total of 933 patients were included in the final meta-analysis. Of these, 671 were from the SANAD cohort (438 cases, 233 controls) and 262 from the AUS cohort (188 cases, 74 controls).

2.5.2. Results of the GWAS

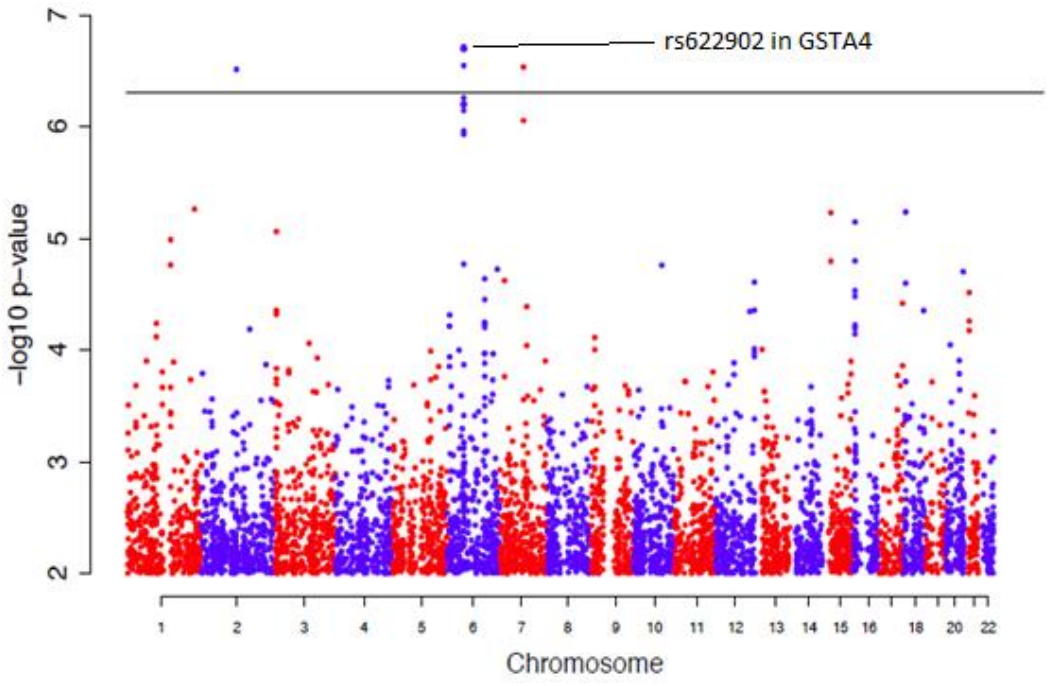

Figure 2.1. GWAS Manhattan plot illustrating the most significant loci of rs622902 ($p = 1.47\times10^{-7}$) in the GSTA4 gene region

The Manhattan plot identified 3 loci with meta p-values $<5\times10^{-7}$ (this significance level is illustrated by the horizontal bar), suggestive of association between the likelihood of 1-year remission and genetic variants in these genomic regions. Of these loci, the strongest

association was observed in chromosome 6; the most significant p-value ($p=1.47\times10^{-7}$, $\log_{10}$ value=6.83) was attributed to the SNP rs622902, located within the gene encoding glutathione-S-transferase alpha 4 (*GSTA4).* Other loci identified included a SNP on chromosome 2 which represents rs13022574 – the DPP10 gene, and rs37919 on chromosome 7 within the region of GNA1, CD36 and GNAT3. The former is a dipeptidyl peptidase-like protein that regulates the functional properties of neuronal Kv4 channels whose implication in epilepsy has been previously described by Maffie and Rudy (13). GNA1 is glucosamine-phosphate N-acetylytransferase; CD36 is a thrombospondin receptor; and GNAT3 is a a guanine nucleotide binding protein. The role of these SNPs on chromosome 7 is less obvious in epileptogenesis.

An overview of GSTA4 is given in the next chapter commensurate with a hypothesis for its role in epileptogenesis and the pharmacoresistant phenotype.

Chapter 3

Glutathione and the glutathione S-transferase enzyme superfamily

Chapter 3: Glutathione and the glutathione S-transferase enzyme superfamily

3.1 Glutathione

Glutathione (GSH) is a molecule of detrimental importance in the physiology of cellular redox balance where, via the action of glutathione *S*-transferase (GST) enzymes, it prevents the oxidation of cellular compounds, organelles and membrane lipids by toxic phase I metabolites. It was discovered by Nobel laureate Sir Frederick Gowland Hopkins of Cambridge University in 1922, where he noted that the addition of GSH could restore a tissues ability to reduce methylene blue, when that tissue had previously lost its reducing power (14).

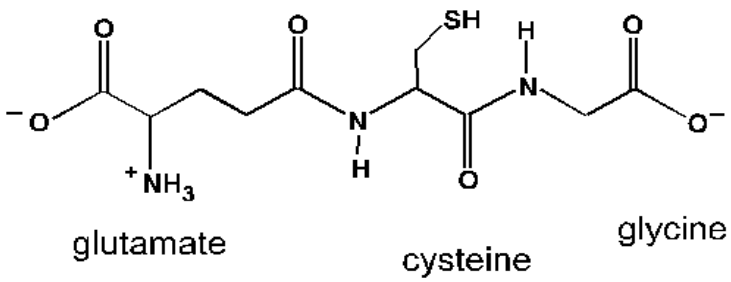

Figure 3.1. The structure of glutathione

Taken from Thomas J, 1999 (15).

Figure 1.2 depicts the structural representation of GSH. It is a water soluble tripeptide antioxidant consisting of glycine, glutamate and cysteine. It is an important molecule, thought to maintain proteins in their reduced thiol state and therefore prevent cysteine residues from becoming oxidised and forming disulphide bonds with one another. Along with its role in the redox reactions of the cell; the role in leukotriene, steroid and prostaglandin synthesis; and its involvement in cell signalling, it has also been reported that GSH may have a part to play in the other basic cell functions such as apoptosis and regulation of the cell cycle (16).

Schafer and Buettner (17) have conducted extensive research into GSH. They have shown that the intracellular concentrations of GSH vary, with the distribution being somewhat compartmentalised. The highest concentration of GSH can be found in the cytoplasm – between 1 and 11mM – mirroring the fact that this is the site of GSH synthesis. The mitochondria also contain a separate pool, where the process of oxidative phosphorylation produces superoxide (O_2^-) that would otherwise destroy the organelle and cell (16, 18). Nuclear GSH maintains the redox state of important sulfhydryls that exists within proteins necessary for DNA repair (18). This compartmentalisation of GSH is important; radical species produced within the cells have a known short half-life, and their subsequent damage occurs close to the sites at which they are produced (16). In order for their effective removal, GSH must be present in these sites of radical manufacture such that the redox balance is maintained.

The common affinity for GSH shown throughout the GST superfamily also illustrates the importance of this molecule as a powerful reducing agent. Whilst the GSTs have affinity for a multitude of substrates and compounds, this affinity shown throughout all members of the cytosolic and mitochondrial GST families for GSH specifically is owed to a GSH binding site in the GST macromolecule – the 'G-site' (19). This will be discussed in greater detail in section 1.2.3.

The synthesis of GSH occurs de novo with the reduced form, GSH, found in much greater concentration than the oxidised form, glutathione disulphide (GSSH). Reduction of GSSH is by the enzyme glutathione reductase (20). The synthesis and degradation of GSH is linked with the γ-glutamyl cycle, which is depicted below in figure 1.3.

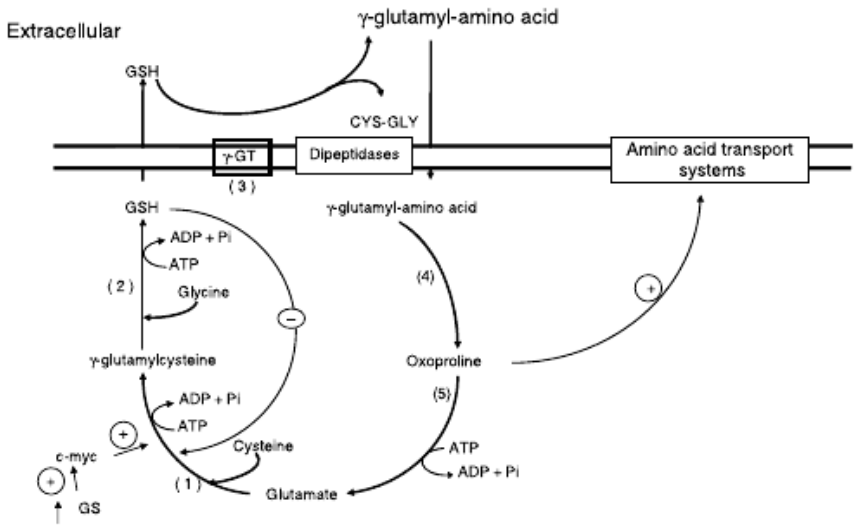

Figure 3.2. The synthesis and degradation of GSH within the γ-glutamyl cycle
Taken from Masella R & Mazza G, 2009 (16).

1) γ-glutamylcysteine synthetase; 2) GSH synthetase; 3) γ-glutamyltranspeptidase; 4) γ-glutamylcyclotransferase; 5) oxoprolinase.

CYS-GLY – cysteinyl glycine; γ-GT – γ-glutamyl transferase

3.2. The glutathione *S*-transferase family

The Glutathione S-Transferases (GST), named thus owing to their ability to transfer a functional group from glutathione (GSH) to an acceptor species, comprises a group of ubiquitous enzymes vital to a plethora of organisms across many kingdoms and phyla (21). Their role is varied, but ultimately they are involved in biochemical reactions that catalyse nucleophylic attack by GSH in the reduced state on non-polar compounds containing an electrophylic carbon, nitrogen, or sulphur atom (19, 22). Such catalysis is detrimental to the function and survival of the cell, as conjugation of GSH with reactive electrophylic intermediates yields compounds that are both less reactive and more soluble (19). Many of the reactive intermediates are by-products of phase I oxidative metabolism and other *in vivo* biochemical reactions (19, 22, 23). However, GSTs astounding affinity for a multitude of substrates enables this group of enzymes to catalyse many xenobiotics and environmental hazardous substances including pesticides, carcinogens, and several

pharmacological agents, and are thereby considered a highly sophisticated evolutionary mechanism for removing cytotoxic agents from within the cell. Further research has also elucidated many other roles, including hormone synthesis, tyrosine degradation, and peroxide breakdown (22). Not only are GSTs involved in enzymatic reactions involving GSH, but they have also been shown to synthesise leukotrienes via mechanisms which utilise GSH as a co-factor, and they are also implicated in the production of prostaglandins (22). These will be discussed in greater detail in the relevant sections.

There exist three superfamilies of GST: cytosolic, mitochondrial, and MAPEG (membrane-associated proteins in eicasanoid and glutathione metabolism), the latter formerly designated microsomal GST (MGST). Both cytosolic and mitochondrial subtypes share deep evolutionary similarities in their three-dimensional structure, but these groups bear no reflection of a common ancestry with the MAPEG enzymes; cytosolic and mitochondrial GSTs form dimers, where MAPEG GSTs form bound trimeric enzymes. Cytosolic GSTs comprise the largest group, and are responsible for the catalysis of the greatest variation of biochemical reactions within the cell. It is, therefore, the group that will be elaborated upon in greatest detail.

As aforementioned, GSTs are also implicated in the detoxification of xenobiotics, and as such these groups of enzymes are of great interest to both pharmacologists and toxicologists. As we learn more regarding the structure, function, and genetics of these enzymes, interest within them from a clinical perspective grows exponentially. GST is involved in the metabolism and detoxification of a multitude of chemotherapeutic agents, with upregulation of these enzymes being found in cancer cells, and henceforth leading to chemotherapeutic resistance (24). Further, their role in leukotriene synthesis has led to a wave of research activity into a target in asthmatic patients, where upregulation of leukotrienes are implicated in bronchiolar constriction. Drugs that antagonise the leukotriene molecules, such as montelukast, are now prescribed.

3.2.1. Nomenclature of the glutathione S-transferases

An internationally accepted method for naming GST was long sought, in part due to the continually evolving knowledge of the function, and diversity, of these enzymes. Upon their initial discovery, GST was known to catalyse the reactions of GSH with bromosulfophthalein (25) and chloronitrobenzenes (26), and subsequently the name GSH S-aryltransferase was given. Other known GSTs involved in the catalysis of epoxides and alkyl halides were given the respective names GSH S-epoxide transferase and GSH S-alkyl transferase (27). However, it was soon discovered that the compound 1-chloro-2,4-dinitrobenzene acted as a 'general' substrate for all GSTs (28), and the original nomenclature had to be abandoned.

There still exists debate in the naming of GSTs across the entire superfamily of enzymes. Decades of research into GST has procured much information on their importance in cellular homeostasis and detoxification of endogenous and xenobiotic compounds, and has thus necessitated a unified nomenclature method. To date, a system of internationally accepted nomenclature exists for the soluble mammalian GSTs, which includes cytosolic and mitochondrial enzymes (the latter of which are also designated GSTK).

The nomenclature of the GSTs is based solely upon amino acid sequence similarity. Enzymes related more closely to others, based on their sequence, are further divided into classes which are identified by the use of the Greek letters: alpha, pi, mu, omega, etc (29). It should be noted that these are abbreviated using the Roman capitals: A, P, M, O, etc, and not their Greek symbols. Among members of the same cytosolic GST class, greater than 40% of the amino acid sequence is identical, whereas less than 25% of the sequence is shared between classes (19, 30). Within a given class, the subunit sequences are given Arabic numerals. When writing the gene encoding a particular GST, *italics* are used; conversely, when referring to the protein, italics are not used. For example, the gene encoding the pi class GST with subunit 1 is written *GSTP1*. As GSTs are dimeric, they

may be homodimeric composed of two copies of subunit 1, or heterodimeric (within the same class), composed of more than 1 subunit. In this case, the homodimeric protein Alpha would be written GSTA1-1; the heterodimeric protein alpha is written GSTA1-2. Similarly, the gene encoding GST class mu subunit 1 is written *GSTM1,* with the homodimeric protein written GSTM1-1. In cases where GSTs belonging to different biological organisms need to be distinguished, a lowercase prefix may be used to denote that species, such as r or p for rat and pig, respectively. In this case, the protein mu class GST with subunit 1 for each would be rGSTM1-1 and pGSTM1-1, respectively.

3.2.2. Basic structural features of the glutathione S-transferases

The first determined structure of any GST was that of the porcine pi subclass, pGSTP1-1, as described by Reinemer et al (31). Since then, the three dimensional structures of many of the other families and classes of GST have been elucidated by x-ray crystallography, including similarities and differences between their protein folding, and the chemical properties by which they interact with their substrates.

Variation exists in the tertiary structure between each of the GST superfamilies and subsequent subclasses; as such, a detailed, intricate analysis of each individual structure is beyond the scope of this text. However, the common structural feature of the GSTs includes an N-terminal thioredoxin-like domain with a C-terminal domain composed of α-helices (19). The sites of binding for GSH and the electrophilic reactive species are denoted the ‘G-site’ and ‘H-site’, respectively (19, 32, 33). The G-site of the dimer enzymes is located within a cleft between the N-terminal of one subunit and the C-terminal domain of the other, and this site is therefore completed only following the dimerisation of the two (32). The N-terminal domain comprises the majority of the G-site, whose topology, along with the H-site, is conserved surprisingly across all GST enzymes (33). The G-site is composed of four β-sheets with 3 adjacent α-helices (‘thioredoxin-

like') (34), while the H-site is constructed of the C-terminal domain, itself α-helical with a contributing loop from the N-terminal domain. Substrate specificity within the H-site is owed to variation in the amino acid sequence at this location (33).

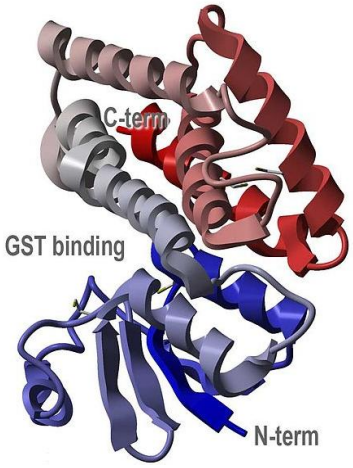

Figure 3.3. The general crystallographic structure of GST illustrating the C- and N-terminus and the common GST binding site for GSH.
Taken from Protein Data Bank:
http://www.rcsb.org/pdb/explore/explore.do?structureId=1R5A

3.2.3. The cytosolic glutathione S-transferases

Cytosolic GSTs are dimeric proteins with between 199-244 amino acids (22). In mammals, there are seven subclasses: alpha (A), mu (M), pi (P), sigma (S), theta (T), omega (O) and zeta (Z). Other subclasses do exist (beta, delta, epsilon, lambda, phi, tau, and "U") (22), though these enzymes are found in non-mammalian species and are not discussed further.

The cytosolic GSTs, besides conjugating reactive species with GSH, are also involved in the intracellular transport mechanisms for the elimination of compounds from within the cell, binding ligands in a non-substrate manner (35). Such compounds include hormones, haem, bilirubin, and steroids, and it was this property that led to their initial name of

Ligandin (35) - this name was changed upon the discovery of the wide functional roles of these enzymes.

The sigma class GST, otherwise known as haematopoeitic prostaglandin D synthase, converts PGH_2 into PGD_2 in a GSH-dependant manner, as outlined in figure 1.5 (36). The theta class GSTs are amongst the least studied of all mammalian GST isoenzymes, in part owing to the fact that they have a reduced affinity for the classical substrate 1-chloro-2,4-dinitrobenzene compared with the other classes (37). Work carried out on the theta class has shown a sulphate-binding active site, which has proposed the hypothesis that these isoenzymes are important in reactions involving sulphate esters and, as such, may be important in the prevention of hepatocellular carcinoma (38, 39). Omega GSTs share common properties with the glutaredoxins, where these enzymes possess a cysteine within their active site and are thus able to form disulphide bonds with GSH and catalyse reductase reactions (40). Both classes mu and pi GST have reported activities that inhibit apoptosis signal-regulating kinase 1 (Ask1) and c-Jun N-terminal kinases (JNK): Ask1 is vital to the mechanisms concerned with cytokine- and stress-induced apoptosis, and its inhibition by these GSTs during non-stressed conditions prevents premature and unnecessary apoptosis (41). JNK, a kinase involved with apoptosis following inflammation, heat shock, UV irradiation and cytokine presence, is particularly inhibited by GST pi, thereby also preventing unnecessary apoptosis (42, 43). Zeta GST, also known as maleylacetoacetate isomerase (MAAI), utilises GSH as a co-factor for the isomerisation reaction of maleylacetoacetate to fumarylacetoacetate (19). This isomerisation is critically important in the metabolism of both phenylalanine and tyrosine (19). Further, GSTZ GSTs is also implicated in reactions involving oxygenation, dehalogenation, and peroxidation, and also exhibits transferase activity - important metabolic clinical syndromes result with enzyme deficiencies of the tyrosine pathway, including phenylketonuria (19). Finally, the alpha class GST is important in the metabolism of a number of environmental carcinogens, including the ubiquitous

polyaromatic hydrocarbons (PAH), and they show particular affinity towards the metabolites of lipid peroxidation (44) and cell membrane oxidation (45).

3.2.4. The mitochondrial glutathione S-transferases

Mitochondrial GSTs, also known as GST kappa (GSTK), share deep structural evolutionary links with the cytosolic group of enzymes (46). They were first isolated from the mitochondria of rat liver by Harris et al (46) and this organelle-specific localisation was confirmed using immunohistochemical techniques by Thomson et al (47) on specimens of liver and kidney. A number of research methodologies by several groups has confirmed the presence of this class of GST in the mitochondria, as well as finding the enzyme within cellular peroxisomes where they act as peroxidase enzymes protecting against lipid-peroxides (30). The specific localisation of this class reflects the fact that GSTK is highly involved in the detoxification of reactive-oxygen species from the respiratory chain reactions and oxidative phosphorylation, particularly more so than other classes (48).

Recently, research undertaken by Liu et al (49) has defined a role for GSTK as an adiponectin-interacting protein. Adiponectin is an adipokine molecule secreted specifically from adipose tissues where it is implicated in the metabolism of both glucose and insulin within insulin-sensitive tissues (50). Several compounds have been shown to be involved in the folding of these adiponectin molecules in the human body, by assisting in the formation of disulphide bonds. Interestingly, it was found by Liu et al (49) that increasing the concentration of GSTK directly correlated with an increase in multimerization of adiponectin by the formation of disulphide bonds, with the converse being true with a decrease in GSTK – this has led to the suggestion that GSTK may be concerned with the correct folding of these important metabolic proteins, and as such expression differences may carry a clinical implication in obesity disorders (51).

3.2.5. Membrane-associated proteins in eicosanoid and glutathione metabolism

The MAPEG series of GSTs contains four subtypes (I-IV), of which there exist six isoenzymes belonging to subtypes I, II and IV in humans. These six isoenzymes are (52):

1) 5-lipooxygenase activating protein (FLAP);
2) Leukotriene (LT) C_4 synthase (LTC_4S);
3) Microsomal glutathione S-transferase 1 (MGST 1);
4) Microsomal glutathione S-transferase 2 (MGST 2);
5) Microsomal glutathione S-transferase 3 (MGST 3), and;
6) Prostaglandin E synthase (PGES).

The initial two isoenzymes are involved in the biosynthesis of leukotrienes; LTC_4S conjugates leukotriene A_4 with GSH, and FLAP is necessary for the production of 5-lipooxygenase from its precursor arachidonic acid. MGSTs 1-3 are GSTs as well as GSH-dependent peroxidases; MGST 1 is implicated in the biochemical reactions conjugating electrophilic compounds with GSH, whilst MGST 2 and 3 both catalyse the conjugation of GSH with leukotriene A_4. Prostaglandin E synthase is involved in the biosynthesis of prostaglandin E_2 (PGE_2) from the precursor molecule prostaglandin H_2 (PGH_2), ultimately responsible for the synthesis of prostaglandin F_2 (PGF_2) - all of these biomolecules are potent mediators of inflammation (53). Figure 1.5 illustrates the synthesis of the eicosanoids.

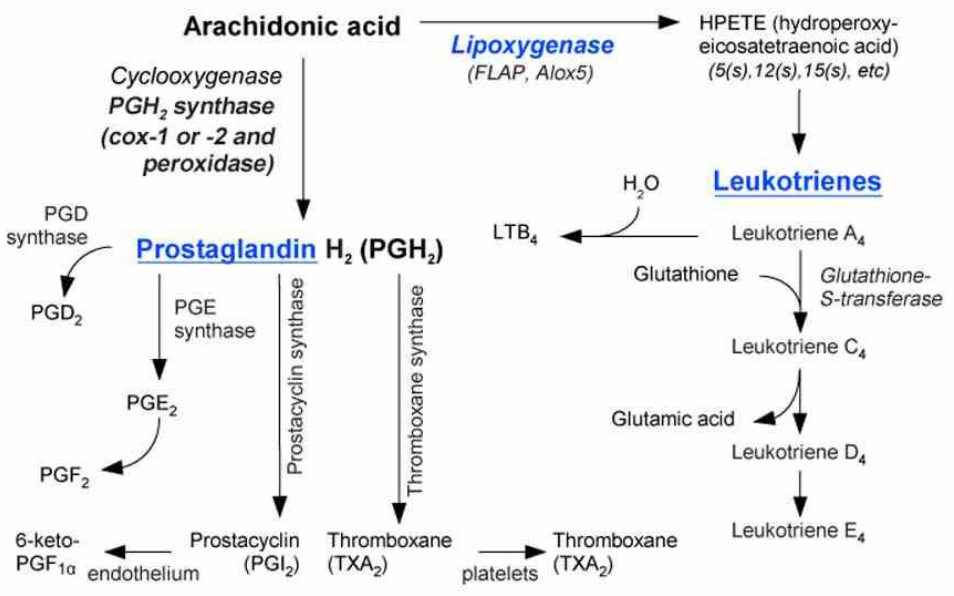

Figure 3.4. The synthesis of the eicosanoids, with the specific role of GSH and GST shown in leukotriene (leukotriene C4) production
Taken from: http://www.scienceofhealthindex.com/images/Eicosanoid_synthesis.jpg

According to subdivision based on sequence, group I MAPEGs contain FLAP, LTC_4S and MGST 2; group II contains MGST 3 as its only member; and group IV contains PGES and MGST 1. Group III subtypes are found only in the bacteria Escherichia coli and Vibrio cholera (22).

Table 3.1. MAPEG subgroup with corresponding isoenzymes and their biological roles

Subgroup	Isoenzyme	Biological Function
I	5-lipoxygenase activating protein (FLAP)	Leukotriene biosynthesis activation
	Leukotriene C_4 Synthase (LC_4S)	GST specific for LTC_4 synthesis
	Microsomal Glutathione S-Transferase 2 (MGST 2)	GST activity & GSH peroxidase activity
II	Microsomal Glutathione S-Transferase 3 (MSGT 3)	GST activity & GSH peroxidase activity
IV	Prostaglandin E Synthase (PGES)	$PGH_2 \rightarrow PGE_2$
	Microsomal Glutathione S-Transferase 1	GST activity & GSH peroxidase activity

3.2.6. Variation within the roles of the cytosolic glutathione S-transferases

A brief overview of variation amongst the human cytosolic GSTs is given below in order to outline their varying roles within the human body, and to give an appreciation of what genetic variants/polymorphisms may have on their function.

For reference, the cytosolic GSTs with their corresponding gene(s) and chromosome number are given below in table 1.4.

Table 3.2. Cytosolic GSTs with known gene(s) and chromosomal number.

GST Class	Known Gene(s)	Chromosome
Alpha	*GSTA1-5*	6
Mu	*GSTM1-5*	1
Pi	*GSTP*	11
Zeta	*GSTZ*	14
Omega	*GSTO1-2*	10
Theta	*GSTT1-2*	22
Sigma	*GSTS*	4

3.2.6.1. Alpha Class

Five GST alphas have been identified to date – GSTA1, A2, A3, A4 and A5 - all located on chromosome 6 (33, 54-56). Using gene-specific probes and real-time PCR methods in 24 human tissue samples, Morel et al (54) showed that GSTA1, GSTA2 and GSTA4 were distributed widely in human tissues, whilst GSTA3 was somewhat rarer and GSTA5 mRNA was not detected. Specifically, transcripts of GSTA4 were detected in all tissues, which was a finding echoed by Desmots et al (57) who found GSTA4 in all tissues by Northern blot analysis. GSTA1 and GSTA2 mRNA were often co-expressed. GSTA3 was found in several of the samples, though in high concentrations in lung, stomach, testis, placenta and suprarenal tissue. In all of the samples analysed, no transcript corresponding to the GSTA5 sequence was detected.

GSTA1 has been implicated in the pathophysiology of inflammation, with its downregulation having profound implications for the mechanisms that guard against

cellular stress from by-products of lipid peroxidation. Ng et al (58) showed that interleukin (IL)-1β, via a mechanism that involved the overexpression of a variant of hepatic nuclear factor 1 (HNF-1), caused the repression of GSTA1 during the inflammatory process. This indicates that GSTA1 may play an important role in the cellular mechanisms that protect against inflammation secondary to the by-products of lipid peroxidation.

The GSTA2 protein is 222 amino acids in length with a molecular mass of 25,425. Its role in health and disease has been studied by several groups; GSTA2 is the known enzyme for conjugation of the myelosuppresive agent busulfan with GSH. This agent may be administered to patients undergoing haematopoietic stem cell transplantation prior to therapy, and its biotransformation appears unaltered by polymorphic variation suggestive of there being no consequence of base variation in these alleles (59). Board and Webb (55) also suggested that this particular GST may be implicated in the aetiology of non-haemolytic unconjugated hyperbilirubinaemia, indicating GSTA2 may have a role in the transport of compounds such as haeme and bilirubin.

GSTA3 is a rarer subclass of the GST alphas and was discovered by Morel et al (54) in their experiments to be a rare message owing to splicing defects. The same group found that the gene contains 7 exons. Ilic et al (60) showed that GSTA3 knock-out mice were particularly sensitive to the cytotoxic effects of aflatoxin B1, whose role is well described in the pathogenesis of hepatocellular carcinoma. The enzyme also has a known role in the biosynthesis of a number of steroids, where it is implicated in the double-bond isomerisation of these molecules (61). It is a rare transcript (62), found largely in tissues where steroid synthesis is high – the ovaries, testes, placenta and adrenals (61). Polymorphism within the GSTA3 gene, therefore, may contribute somewhat to the pathology within steroid-producing tissue, though this link has not been investigated to date.

GSTA4 has a wide tissue distribution in humans. Amongst other organs, it is found within kidney, liver, brain, skin, heart and colon (63). This particular alpha isoform was isolated and cloned from a human cDNA library (62), whereupon it was shown to have particular enzymatic activity towards carbonyl compounds and reactive radical species (63), indicating a significant role for this isoform in cellular homeostasis. The role of GSTA4 will be discussed in further detail at the end of this section.

To date, little is known regarding the role of GSTA5. Hayes et al (64) reported the upregulation of GSTA5 in rat hepatocytes following the administration of chemotherapeutic agents, with further computer modelling of the promoter region of *GSTA5* identifying a possible antioxidant response element that may confer this resistance (64, 65). The postulated role of this enzyme in chemoresistance suggests their upregulation may also play a role in other diseases where pharmacoresistance is a problem, though this hypothesis has yet to be explored.

3.2.6.2. Mu class

The five known GST mu genes, *GSTM1-5*, have been mapped to chromosome 1 using locus-specific PCR primer pairs on DNA samples retrieved from human/hamster somatic cell hybrids (66). Of the GSTM subclass, both GSTM1 and GSTM2 have been shown to be the most active GSTs within the liver (67). The specific roles of these isozymes within the hepatic parenchyma, however, are poorly understood.

Homozygosity for the *GSTM1* null allele has been linked with carcinomas of the head and neck as discovered by Lohmueller et al (68) in a large metanalysis of 301 published genetic association studies. Another study by Huang et al (69) initially identified a statistical association between the SNP rs366631 and GSTM1 expression using GWAS. This SNP is found approximately 11kb downstream of the *GSTM1* gene in the 3' region. Later, the association was found to be invalid, and the false genotyping call arose from sequence homology and a *GSTM1* region deletion. Following their work, the group

concluded that rs366631 is a pseudo-SNP and may, perhaps, be used as a biomarker for *GSTM1* gene deletion. *GSTM1* null has been implicated in several disease states, and its role in epilepsy is discussed further at the end of this chapter.

In comparison to GSTM1, much less is known regarding the function of GSTM2, M3, M4 and M5. GSTM2 is a dimeric enzyme that is found mainly within liver and muscle tissues, and GSTM3 is found largely within the brain (70). Studies into the significance of *GSTM4* gene variation in human disease are few, though one group suggested that a T2517C polymorphism within intron 6 of the *GSTM4* gene is associated with carcinoma of the lung (Fisher's Exact Test, p=0.026) in a population containing 136 lung cancer patients and 156 healthy controls (71).

GSTM5 was discovered by Takahashi et al (72) by screening a human frontal cortex cDNA library with a rat cDNA that cross-hybridised with other human and rodent GSTM cDNAs. This isozyme was found to be expressed primarily in brain and lung tissue by Northern blot analysis, and is predicted to contain 217 amino acids. The role of GSTM5 has yet to be outlined.

3.2.6.3. Omega class

Two members of the omega class are known to exist to date, whose loci exist on chromosome 10: GSTO1 and GSTO2 (40, 73). Whilst all of the GST isozymes share structural and functional variance, the GST omega is somewhat further distinct from the others (33).

GSTO1 cDNA was discovered by searching an expressed sequence tag (EST) database for homologies with *GSTZ1*. It is most closely related to GST zeta and contains 241 amino acids (40). Whitbread et al (73) illustrated the ability of GTSO1 to reduce monomethylarsonic acid to monomethylarsonous acid, which is the rate-limiting step in the metabolism of inorganic arsenic. This class of GST possesses a 19 residue N-terminal extension forming a novel structural unit whose function remains elusive, and that is not

seen within any other class of GST. Further, substrates that are biochemically altered by other GSTs are not altered by certain *GSTO1* alleles, and its ability to metabolise dehydroascorbate in a GSH-dependant reduction shows that this particular allele illustrates behaviour that is characteristic of the glutaredoxins rather than the GSTs (40, 74).

The role of GSTO2 currently remains unclear. Significant expression was found in the testes, though amounts have also been shown in kidney, liver, striated muscle and prostate tissues (73). This allele shares 64% of its amino acid sequence with *GSTO1* (73).

3.2.6.4. Pi class

One member of the GST pi is currently known to date, and is coded by a loci found on chromosome 11 (33). Its expression can be found in all tissues of the human body, expect for red blood corpuscles and little amounts found in the liver (70, 75).

GSTP was isolated by Kano et al (76) by screening a human placental cDNA library with rat placental GST cDNA. The protein is approximately 209 amino acids in length, and contains 7 exons which span approximately 2.8kb (77).

Several studies have suggested a role for GSTP in a number of biochemical pathways, and its role in several disease states has been implicated. Expression of GSTP was found to be increased in many tumours relative to normal tissue samples, suggesting a role for GSTP in carcinogenesis (78). These links are poorly explored.

A role in hepatotoxicity following overdose of paracetomol has also been outlined in experiments using GSTP. Mice null for *GSTP* were examined for the metabolic effects of this genotype, and it was discovered that these mice were highly resistant to the effects of the toxic paracetomol metabolite, NAPQI (79). The results of this analysis suggest that variation in *GSTP* may contribute to the toxicity following overdose of paracetomol.

Much interest in the polymorphism within *GSTP* has been shown. Of particular interest was a study conducted by Menegon et al (80), who pursued the widely-accepted hypothesis that Parkinson's disease could be caused by numerous environmental factors, including xenobiotic agents and pesticides. Given the role of GST in the reduction of such compounds, and its presence in the blood-brain barrier (81), the group investigated the role of GST polymorphisms in the development of Parkinson's disease in 95 patients and 95 controls by PCR genotyping. When analysing the results of those patients who had reported pesticide exposure, associations were seen only within GSTP, with differences secondary to an excess of heterozygotes and non-carries of the A allele in the patient group compared to the control group.

3.2.6.5. Theta class

Two isozymes comprise the GST theta class (GSTT1 and GSTT2), whose location on chromosome 22 was found by in-situ hybridisation studies (82). Pemble et al (83) are responsible for the cDNA cloning of this class, and deduced its composition of 239 amino acids, though for some time GSTT was overlooked owing to its low activity with the classic GST substrate, 1-chloro-2,4-dinitrobenzene.

A study by Chen et al (84) compared the frequency of the *GSTT1* null genotype in 96 myelodysplastic syndromes and 201 healthy controls matched for age, race and sex. Forty-six per cent of patients with myelodysplastic syndrome were null for *GSTT1* compared with 16% of controls. The results illustrated a 4.3-fold increased likelihood of myelodysplastic syndrome, though the mechanism of this increased inference was not understood.

GSTT2 contains 244 amino acids and has a 55% sequence identity to GSTT1 (85). Little is understood regarding the role of this isozyme, though it has been suggested that SNPs in the promoter region of this gene (-537G>A, -277T>C and -158G>A) may confer greater risk of colorectal cancer (86).

3.2.6.6. Zeta class

GST zeta was discovered using sequence alignment and phylogenetic analysis, and is located on chromosome 14 (87, 88). This isoenzyme is found in abundance in the liver and proximal convoluted tubules of the kidney, where they catalyse GSH-dependant reactions involving an array of electrophilic alpha-halogenated acids (33). GST zeta is also characterised as maleylacetoacetate isomerase owing to its role in the catabolic pathway of both phenylalanine and tyrosine (89).

Polymorphisms within *GSTZ* are known to have clinical consequence. Decreased ability to catalyse the aforementioned reactions lead to metabolic diseases including type I tyrosinaemia and phenylketonuria (90).

3.2.7. An overview of glutathione *S*-transferase alpha 4

Board et al (62) was responsible for isolating the entire GSTA4 cDNA clone from a human adult brain cDNA library. The protein is 222 amino acid molecules long with a molecular mass of 25.7kD. It shares 52% sequence identity with GSTA1, spans 18kb, and contains 7 exons.

Within the *GSTA4* gene region there are, to date, a total of 361 known SNPs (contig label: GRCh37.p5) including the intronic regions. Of these, 12 are known functional SNPs (non-synonymous). Fourteen exist in the 3' UTR region and 5 are within the 5' UTR region of the gene. The chromosomal coordinates are 6p12.1 and it can be found within base pairs 52,842,746 – 52,860,178 (91).

The GSTA4 enzyme has been shown to have two specific functions. The enzyme shows a high affinity for 4-hydroxynonenal (4-HNE), a metabolic by-product of peroxide degradation from lipid membranes and arachidonic acid (92). At times of increased cellular damage, inflammation and metabolic stress, levels of 4-HNE rise as a result of peroxide degradation, and it is therefore thought that the role of GSTA4 in preventing this damage is important. Secondly, Board (62) showed that GSTA4 shows high affinity for

the catalysis of reactive carbonyl compounds such as alk-2-enals whose presence also lead to cellular stress and damage. Combined, the role of GSTA4 as a protector against reactive intermediates is considerable.

A number of studies have been conducted into the link between GSTA4 and disease. It is known that insulin resistance is linked to an increase in the production of free radical species, resulting in the increased production of lipid aldehydes (93). These aldehydes are enzymatically degraded primarily by GSTA4. Decreased levels of GSTA4 and resultant increases in lipid aldehyde production were shown to cause abnormal protein carbonylation within adipocytes. The effect was aberrant glucose transport, lipolysis and mitochondrial respiration. Downregulation of GSTA4 may, therefore, contribute to the pathogenesis of insulin resistance and type 2 diabetes. Its role in other diseases may also be significant.

Polymorphisms within *GSTA4* have been associated with carcinoma of the lung. A genetic study in a Southeastern Chinese cohort by Qian et al (94) genotyped 3 SNPs (rs182623 – 1718:T>A, rs3798804 + 5034:G>A and rs316141 + 13984: C>T) in 500 lung cancer patients and 517 healthy controls. A significant association was observed in genotype and allele frequency distributions in -1718 between cases and controls (p=0.006 and p=0.003, respectively). When comparing with the -1718TT genotype, those participants with the TA and AA genotypes were at a decreased risk of developing carcinoma of the lung (OR 0.63, 95% CI: 0.47 – 0.84, p=0.006).

3.2.8. The glutathione S-transferases and their role in epilepsy and antiepileptic drugs: Current literature

The literature regarding the role of GSTs in epilepsy and its treatment is scarce, with very few published studies. There are, however, a small number of groups that have explored the role of certain cytosolic GST subclasses in epilepsy, or in their effect on AEDs.

Liu and Tsai (95) studied the effects of the null genotype GST Mu (*GSTM1(-)*) in a cohort of 1323 patients; 97 patients with epilepsy with good seizure control on anticonvulsant therapy, 35 patients with intractable seizures on anticonvulsant polytherapy, and 1191 healthy control subjects. Good seizure control was defined as fewer than 2 seizures a month with monotherapy, and intractable seizures were defined as a seizure frequency of greater than 4 per a month whilst taking a combination of anticonvulsants with effective monitoring of blood levels for at least four months. Patients with generalised tonic-clonic seizures or secondary generalised tonic-clonic seizures were included in the study, and the pharmaceutical agents prescribed, either alone or in combination, included phenytoin, carbamazepine, valproic acid, and phenobarbital. Both null genotype of GSTM1 and the plasma level of malondialdehyde (MDA) were determined for the groups of patients outlined above. MDA is a marker of oxidative stress, and was used to indirectly measure the level of lipid peroxidation in the patients. The results showed that the allele frequency of GSTM1(-) genotype of the patients with intractable epilepsy was significantly higher than those with well-controlled epilepsy (80% vs. 48.5%, OR = 4.3, p = 0.001) and healthy controls (80% vs. 49.5%, OR = 4.1, p = 0.001). It was therefore suggested that the null Mu genotype may be a significant genetic factor that contributes to the highly variable response to anticonvulsant therapy observed in clinical practise.

The *GSTM1(-)* genotype was also the topic of research, along with the GST theta null genotype *(GSTT1(-)),* for the hypothesis that either of these genotypes alone, or in combination, could be a risk factor for carbamazepine-induced mild hepatotoxicity (96).

GSTM1(-) and *GSTT1(-)* were chosen due to their association with troglitazone-induced liver toxicity, a drug of the thiazolidinedione class used as an anti-glycemic agent whose licence was withdrawn following idiosyncratic hepatotoxicity. The study followed 192 Japanese epilepsy patients treated with carbamazepine, and the genotypes of *GSTM1(-)*, *GSTT1(-)*, and of microsomal epoxide hydrolase-3 and 4 (mEH-3, mEH-4), were determined. mEH is an enzyme involved in a wide range of reactions involving epoxide intermediates, where it is responsible for the addition of water to these substrates. The proportion *GSTM1* null and *GSTT1* null genotypes in their cohort were reported as 55.7% and 39.6%, respectively. In patients with these genotypes, the alanine aminotransferase (ALT) and aspartate aminotransferase (AST) were abnormally elevated in 24% of the *GSTM1* null patients, and 32.3% of the *GSTT1* null patients, with levels rose by approximately 2.3- and 1.8-times that of the normal limit. When comparing both ALT and AST levels in the *GSTM1* null vs. *GSTM1* present genotypes, the levels were significantly higher in the former (ALT - $p = 0.007$ and AST – $p = 0.004$). The levels of ALT was significantly higher in *GSTM1-/GSTT1-* than in *GSTM1+/GSTT1-* and *GSTM1+/GSTT1+* ($p = 0.01$ and $p = 0.01$, respectively). Serum levels of AST was significantly higher in both *GSTM1-/GSTT1-* and *GSTM1-/GSTT1+* than they were in *GSTM1+/GSTT1+* ($p = 0.02$ and $p = 0.003$, respectively). Genotype of mEH showed no significant association with hepatotoxicity. In summary, the group suggested that the *GSTM1* null genotype may be significantly associated with carbamazepine hepatotoxicity.

Interestingly, a similar study was undertaken by Fukushima et al (97) who looked into the effects of the mu null genotype either alone, or in combination with, the theta null genotype on mild hepatotoxicity in patients treated with VPA. Hepatotoxicity was defined as a rise above normal levels of ALT, AST, γ-glutamyltranspeptidase (GGT) and total bilirubin in 149 Japanese patients with a diagnosis of epilepsy. The odds ratio (OR) of the *GSTM1-* vs. *GSTT1+* genotype and the *GSTM1-/GSTT1-* vs. *GSTM1+/GSTT1+*

genotypes for the increase of GGT over the normal limit were 2.3 (95% CI 1.1 – 7.2) and 6.5 (95% CI 1.5 – 28), respectively. Liver function tests were not significantly affected by the *GSTT1* genotype alone. In patients treated with VPA for greater than 6 months, ALT, AST and GGT were all significantly raised in the *GSTM1-* genotype compared to the *GSTM1+* genotype. The authors therefore concluded that both *GSTM1-* and *GSTM1-/GSTT1-* genotypes may be genetic risk factors for the particular increase of GGT in epilepsy patients treated with VPA.

In a research project conducted by Shang et al (81), 32 clinical cortex specimens were obtained following neurosurgical removal for intractable epilepsy. Several structural pathologies were seen, including hippocampal sclerosis, arteriovenous malformation, old infarct, and viral encephalitis. Eight non-epileptic patients who had undergone neurosurgery for arteriovenous malformations for reasons other than epilepsy were used as controls. Patients had been treated for over two years with at least 3 antiepileptic drugs, including phenytoin, phenobarbital, carbamazepine, valproate and topirimate to no effect, with patients experiencing at least 3 seizures per month in the preceding 6 months. Immunohistochemical techniques were employed to detect three GST isoforms in the cortex specimens: alpha, mu and pi. GSTA was not found in any specimens; GSTM was found in 63% of controls and 53% of intractable epileptic specimens; and GSTP showed immunoreactivity in 66% of the epilepsy patients compared with 50% of the controls ($p < 0.01$). These results suggest that high levels of GST pi may contribute to refractory epilepsy.

Whilst the pathophysiology of epilepsy is poorly understood, it is clear that there exists abnormal neuronal cellular function within the central nervous system. Further, in the pharmacoresistant phenotype, there may be local metabolic reasons as to why antiepileptic medications fail to provide adequate control of seizures. With the role of GSH and GST well-described in the homeostasis of cellular redox balance and in the elimination of xenobiotics, it is plausible that polymorphic variation within GST may

contribute to neuronal dysfunction that manifests as epilepsy, or indeed local increased metabolic activity may interfere with antiepileptic drug metabolism. A recent meta-analysis of two genome-wide association studies highlighted SNP variants in GSTA4 to be associated with pharmacoresistant epilepsy.

3.3. Pharmacogenetics and single nucleotide polymorphisms

Pharmacogenetics comprises the study of how the pharmacokinetic and pharmacodynamic factors of a drug vary between individuals owing to genetic differences in the population. It is thought that polymorphisms in drug transporters, channels, metabolising enzymes, etc, may explain to some degree the dose-response variation seen in clinical medicine, and may indeed explain the pharmacoresistant phenotype observed in several diseases (98). The ultimate goal of this area of science is personalised medicine – genotyping of individuals in order to predict their response to a drug, or predict which members of the population may react adversely to a particular drug.

Since the advent of pharmacogenetics, many drugs, their target sites, and metabolising enzymes, have been extensively researched. A particular example is outlined with warfarin, a vitamin-K antagonist used widely as an anticoagulant for various indications. In dosing a patient with warfarin, only the weight of the patient is taken into consideration, with no other predictors of how well the patient will respond taken into account. Maintaining serum concentrations within the therapeutic index requires constant monitoring and dose adjustment owing to inter-individual drug response, and it is thought this variation may be secondary to genetic differences, particularly polymorphisms in the genes *VKORC1* and *CYP2C9* (99-101). In the United States, genetic testing is now indicated in order to better predict response to warfarin, and it is hoped that this pre-prescription testing many occur for more drugs in the future (100, 102).

Single nucleotide polymorphisms (SNPs) are variations within a single nucleotide, such that one base is switched for an alternate. In describing these variations, we refer to a SNP as having two alleles, such that a change from a cytosine (C) to thymine (T) (C>T) SNP would have C and T alleles, and the individual could be genotype CC, CT or TT. SNP variation is the commonest form of genetic variation, and there are thought to be in excess of 3 million common SNPs within the human population (12). The vast majority of these polymorphisms are in regions of the genome that have no functional significance, with protein function remaining unaltered. In order of perceived functional significance, there major categories of SNP are:

1) Non-synonymous
    a. Nonsense
    b. Missense
2) Splice site
    a. 5' splice site – upstream of the gene
    b. 3' splice site – downstream of the gene
3) mRNA untranslated region (mRNA UTR)
    a. 5' UTR – upstream of the gene
    b. 3' UTR – downstream of the gene
4) Synonymous
5) Intronic

SNPs within each of the above genomic regions may affect protein structure and, hence, may affect function in a variety of ways. Nonsense SNPs generally affect protein function most drastically, where they are responsible for premature stop-codons and the truncation of the polypeptide. It is also known that nonsense mutation can result in mRNA decay, resulting in the complete absence of the gene product – nonsense-mediated decay (103).

Missense SNPs are responsible for the change in the amino acid that occur secondary to an alternate codon being produced. The resultant change may or may not result in a functional difference in the protein, as the substituted amino acid may or may not cause a significant change.

The process whereby intronic regions of DNA are removed and exonic regions are joined is termed splicing. SNPs within the splicing sites (either up- or downstream) of the gene thereby result in defected amino acid sequence and may impact upon the specificity of the transcribed protein.

SNPs within the mRNA UTR regions can also occur either up- or downstream of the gene. These regions of DNA are implicated in the binding of transcription factors and in the stability of the mRNA, and thereby variation within these sites may affect protein synthesis and overall protein structure in these ways.

Synonymous SNPs result in codon-triplets that code for the same amino acid as previously, and it was therefore first thought that these genetic variants were of no functional consequence. However, it is now recognised that certain codons translate for amino acids more readily and efficiently than others, and therefore these SNPs may also have a consequence for the overall gene product if the polymorphism is at a critical region.

Those SNPs that exist in the non-coding, intronic region of the gene are thought to carry the least functional significance. Whilst these variants are not directly associated with the overall three-dimensional structure of the protein, they may have a role in the stability of the gene and in gene expression during the transcription-translation process.

3.4. Aims

The aim of this study was to further investigate the association of *GSTA4* gene variants with treatment outcome in newly-diagnosed epilepsy. This was undertaken by detailed analysis of multiple polymorphic sites across the *GSTA4* gene in an effort to validate the original observation, and to identify the potentially causal variant. The analysis was achieved by selecting 17 SNPs from across the *GSTA4* gene that represented variation at a total of 48 loci, genotyping those SNPs using both TaqMan real-time PCR and Sequenom MALDI-TOF MS methods, and performing binary logistic regression and Cox multivariate regression analysis of SNP genotype against treatment response with adjustment for clinical covariates as described in the original GWAS analysis.

Chapter 4

Methods

Chapter 4: Methods

4.1. SNP selection

4.1.1. HapMap

The HapMap project is an international collaboration between six countries whose aim is to produce a publically accessible database of the known genetic variation within Homo sapiens in order to advance the research of medical genetics and their effects on disease.

The HapMap database (accessible via: http://hapmap.ncbi.nlm.nih.gov/) was used to obtain the chromosomal coordinates of the *GSTA4* gene using the genome browser #27 feature (details given below). Ten kilobases (kb) of DNA sequence were added up- and down-stream of these coordinates such that the 5' upstream and 3' downstream regions of the gene were confidently included in the search.

Table 4.1. Chromosomal coordinates of the GSTA4 gene

Gene	Chromosomal position (kb)	Expanded position (±10kb)
GSTA4	Chr 6: 52,950 – 52,968	Chr 6: 52,940 – 52,978

4.1.2. HaploView

HaploView (version 4.2) is a bioinformatics software package that allows linkage disequilibrium (LD) and haplotype block analysis, haplotype population frequency estimation, and SNP and haplotype association testing amongst its features. The coordinates obtained from the HapMap database were entered into HaploView and a comprehensive list of SNPs representing variation across the *GSTA4* gene region (± 10kb) in the Caucasian only population was generated.

This list was refined within HaploView by applying the following restrictions and rescoring the genetic markers:

- Hardy-Weinberg (HW) p-value = 0.001
- Minimum genotype: 75%
- Maximum number Mendelian errors: 1
- Minimum minor allele frequency: 0.001

This resulted in the identification of 48 SNPs representative of variation across the *GSTA4* gene region. Thereafter, the 'Tagger' function within HaploView was employed to generate a list of tagging SNPs that, owing to the LD structure across the *GSTA4* gene region, would serve as markers for these 48 alleles and thereby reduce the number of SNPs required to genotype. The following criteria were set for the identification of tagging SNPs in Tagger:

- Pairwise tagging only
- r^2 value = 0.8
- LOD threshold for multi-marker tests = 3.0

The resulting list of tagging SNPs was manipulated to include SNPs of specific interest (force-inclusion). These are listed in table 3.2 below. Force inclusions included SNPs typed in the original GWAS, those for which evidence of association existed, and those in putatively functional locations within the gene (i.e., coding regions, splice sites, promoter region). The latter characteristics were determined by reference to the National Centre for Biotechnology Information (NCBI) website (accessible via: http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp).

Table 4.2. List of forced inclusion SNPs

SNP	Reason
rs384505	Reported as significant in AUS cohort
rs622902	Typed in GWAS
rs4147617	Potential missense SNP, though with very low MAF
rs316132	Typed in GWAS
rs1802061	Synonymous SNP

This resulted in the final identification of 17 SNPs across the *GSTA4* gene region that would effectively (r^2=0.978) capture all known genetic variation. Those tagging SNPs are listed in table 3.3, together with the alleles captured by each SNP.

Table 4.3. Tagging/test SNPs with captured alleles

Test	Captured alleles
rs384505	rs384505
rs622902	rs622902
rs4147617	rs4147617
rs316132	rs612483, rs426013, rs584124, rs316133, rs670960, rs316141, rs316129, rs316132, rs316140, rs613764, rs316131
rs1802061	rs17615213, rs1802061, rs4986947, rs17614751
rs316128	rs375872, rs316135, rs375887, rs428957, rs449690, rs385636, rs316128, rs419129, rs426169, rs387853
rs13191212	rs672822, rs669674, rs654144, rs11967816, rs13191212, rs508078
rs6922246	rs6922246, rs1032419, rs9463851
rs34511626	rs13219629, rs34511626
rs6904769	rs6904769, rs7496
rs3734431	rs3734431
rs16883343	rs16883343
rs17608022	rs17608022
rs6904771	rs6904771
rs396216	rs396216
rs3756980	rs3756980
rs2397136	rs2397136

4.2. Preparing a working-stock of DNA

Prior to plating out the DNA, each of the SANAD and Glasgow samples was Nanodroped® in order to obtain their original concentration. Thereafter, nanopure water was added to the stock DNA samples accordingly in order to obtain working-stock aliquots of 20µL at a concentration of 20ng/µL.

For stock concentrations that were too low, or for those samples that contained insufficient volume, these patients were excluded from the original eligible list. This resulted in 35 stock samples that were unusable from the SANAD cohort. All samples from the Glasgow cohort were usable.

The working-stock DNA was plated out into 96-well plates. For purposes of quality control, wells B3 and D7 were filled with nanopure water (blanks), and 8 controls taken from random wells were dispensed into wells H8 – H12 such that following the genotyping the results could be checked for concordance. The plates were then stored at -20^{o}C until required.

4.3. TaqMan® genotyping

PCR is a molecular technique used to amplify a specific segment of DNA across several orders of magnitude such that there are sufficient copies to be analysed. The technology is now extremely commonplace, and its uses include amplification of nucleic acids in medical diagnostics, genotyping and forensic science. TaqMan® PCR is one of several types of 'real-time' PCR methods, so called owing to the ability to quantify, in real-time, the absolute increase in amplified DNA as the reaction proceeds. The key difference in this method is the employment of a TaqMan® probe. The probe is an oligonucleotide and contains a reporter flurophore covalently attached to the 5'-end of the molecule. At the 3'-end is a covalently-bonded quencher molecule. Whilst the two molecules are attached to the probe – prior to the action of Taq polymerase – the quencher prevents fluorescence

from the reporter molecule via fluorescence resonance energy transfer and therefore inhibits detection by the PCR detector. The probes are designed such that they anneal to a region of DNA that is amplified by a specific set of primers. When the Taq polymerase adds nucleotides to the template DNA, the reporter molecule is separated and is free to emit fluorescence that can be detected. Therefore, the amount of fluorescence emitted increases exponentially commensurate with the quantity of DNA and can be reported in real-time.

4.3.1. Plating DNA for PCR genotyping

The stored working-stock DNA was allowed to thaw from -20^{o}C at room temperature and centrifuged at 1,500rpm for 30 seconds. Using an 8-channel multi-pipette, 1µL (20ng) of working-stock DNA was transferred into the alternate wells of a 384-well plate. In this way, four 96-well plates were transferred into one 384-well plate, and then placed in a vacuum centrifuge at room temperature for 5 minutes until the aqueous supernatant had evaporated and the dried DNA was left. If required, this could be sealed and stored at -20^{o}C for future use.

4.3.2. Preparation of polymerase chain reaction solution

For each full, and each half, 384-well plate (depending on how many samples were being genotyped), the quantities shown in table 3.4 were prepared in either a 5mL bijou tube, or a 1mL eppendoft tube, again depending on the quantity required.

Table 4.4. Reagents and volumes for PCR mix

Reagent	Storage (temperature)	384 plate Vol. (µL)	192 plate Vol. (µL)	Storage Temp (^{O}C)
Nanopure water	Sterile bijou tube (room)	950	475	Room
TaqMan master mix	PCR fridge (5^{o}C)	1,000	500	-20
SNP	PCR freezer (-20^{o}C)	50	25	-20
Total		**2,000**	**1,000**	

The SNP was removed from the freezer and allowed to thaw in the PCR room at room temperature. Once thawed, the reagents detailed above were pipetted into a bijou tube under a PCR-only fume cupboard and then mixed using a whirlmixer® for 10seconds. Into a 96-well 'v-plate', 250µL of the PCR mix was pipetted into each well in the first column, A1 – H1. An 8-tip multi-channel pipette was then employed in order to pipette 5µL of the PCR mixture into each of the 384 wells containing the dehydrated DNA. The plate was sealed with a clear adhesive PCR LASER cover thereafter, mixed using a whirlmixer® for 30seconds, and centrifuged at 3,000rpm for one minute.

4.3.3. PCR reaction

The sealed plate was placed in the automated draw of the TaqMan PCR machine and the software set-up and run as per the manufacturer's instructions.

4.4. Sequenom MALDI-TOF genotyping

Fourteen of the seventeen SNPs were genotyped using the Sequenom platform, a matrix assisted laser desorption/ionisation time of flight mass spectrometry technology (MALDI-TOF MS). Using the MySequenom® online gene tool facility (accessible via: https://www.mysequenom.com/Home; login details required) the experimental design was constructed by Sequenom in the manner detailed below.

4.4.1. MySequenom

The 'MySequenom' facility enables easy online assay design using a 5 step tool. The steps involved include:

1. Retrieve and format sequences
   - This step involves the uploading of the SNP (by rs identification) from the NCBI database as described previously.
2. Find proximal SNPs (ProxSNPs)
   - SNP sequences are then aligned to the specified genome followed by re-formatting with registered SNPs in close proximity to the desired assay.
3. Identification of optimal primer areas
   - Next, reformatting of the SNP sequences in order to demark PCR amplification primers is carried out, which will result in a unique amplification product.
4. Assay design
   - The final assay design is produced at this stage.
5. Validation (pleXTEND)
   - The final step in the process is the validation step, whereby the software analyses the final assay design in order to identify any unintended cross-amplification products that may lead to false positive signalling in the mass spectrometer.
   -

4.4.2. Plating DNA for Sequenom Genotyping

DNA from the working-stock, stored at -20^{o}C, was allowed to thaw at room temperature. Two microliters (40ng) was pipetted using an 8-tip multi-channel pipette into a 96-well PCR plate and dried down using a vacuum centrifuge at room temperature for 10 minutes.

4.4.3. Preparation of plex primer mix

For each of the 14 SNPs, the following primers were added into a 1,000 µL eppendorf tube to produce a solution containing 500µL at 0.5µM. This was then placed in the 5^{o}C PCR fridge until required:

- 5µL of forward primer (x14)
- 5µL of reverse primer (x14)
- 360µL nanopure water

Next, the PCR reaction mixture was made under a fume cupboard in the PCR room according to table 2.4.

Table 4.5 Sequenom PCR preparation mix

Reagent	**Volume for Single Reaction**	**Volume for 24-plex (384-wells)**	**Storage Temp (^{o}C)**
Nanopure water	2.85 µL	1368.00 µL	Room
PCR buffer (10x)	0.625 µL	300.00 µL	-20
$MgCl_2$ (25mM)	0.325 µL	156.00 µL	-20
dNTP mix (25mM)	0.10 µL	48.00 µL	-20
Primer mix (0.5µM)	1.00 µL	480.00 µL	-20
Hot Star Taq (5U/µL)	0.10 µL	48.00 µL	-20
Total Volume	**5.00 µL**	**2400.00 µL**	

The reagents were stored according to the temperatures detailed above, and allowed to thaw at room temperature. They were then mixed in a sterile 5,000µL bijou tube. Thereafter, the solution was mixed using a whirlmixer® for 30 seconds, and then stored on ice to prevent action of the Taq enzyme.

Using a 96-well 'v-plate', 300µL of the PCR mix was pipetted into wells A1 – H1. An 8-tip multi-channel pipette was then used to transfer 5µL of this reaction mix into each of

the 384 wells of the dehydrated DNA. The plate was sealed with an adhesive PCR plate seal, whirlmixed® for 1 minute, and centrifuged for 1 minute at 2,000rpm.

4.4.4. PCR Reaction

The initial PCR reaction was run using a PCR cycler as per the manufacturer's instructions.

4.4.5. Shrimp alkaline phosphatase

It is necessary, prior to the addition of further reaction agents in subsequent stages, to add shrimp alkaline phosphatase (SAP) enzyme to the previous PCR products in order to prevent the interference of unincorporated deoxyribonucleotide triphosphate (dNTP). The SAP enzyme results in the dephosphorylation of the excess dNTP and renders it inactive for further reactions.

4.4.5.1. Preparation of the SAP reagents

The SAP enzyme solution was prepared according to the volumes given below.

Table 4.6. SAP Enzyme Reagent Preparation

Reagent	Volume for Single Reaction	Volume for 384-wells	Storage Temp (^{o}C)
Nanopure water	1.53 µL	810.78 µL	Room
hME Buffer (x10)	0.17 µL	90.08 µL	-20
Shrimp Alkaline Phosphatase	0.30 µL	158.98 µL	-20
Total Volume	**2.00 µL**	**1059.84 µL**	

The SAP reagents were taken from the freezer and allowed to thaw at room temperature. Once thawed, the reagents were added into a sterile 5,000µL bijou tube according to the quantities detailed above. The solution was mixed by a whirlmixer® for 30 seconds, and stored on ice to prevent any action or denaturation of the enzyme during further plate preparation.

Eighty-five microliters of the SAP enzyme mixture was then pipetted into wells H1 – H12 of a 96-well 'v-plate'. An 8-tip multi-channel pipette was then used to dispense 10µL from row H into each well in rows A – G.

4.4.5.2. Addition of SAP enzyme to DNA-containing plate

Prior to the addition of the SAP enzyme, both the 96-well SAP-containing plate, and the 384-well PCR plate, were sealed and centrifuged for 1 minute at 2,000rpm.

Using the automated Matrix Liquid Handler (MLH), 2µL of SAP enzyme mixture was added to each of the 384 DNA-containing wells containing the previous PCR reaction products. The software protocol, produced by Sequenom for the use of this machine, was consulted in order to operate the MHL.

4.4.5.3. SAP enzyme reaction

The SAP reaction was undertaken on a PCR cycler as per the manufacturer's instructions.

4.4.6. iPLEX® gold reaction mixture

Details of the mass of each extension primer were supplied by the manufacturers in a table that could be utilised for the preparation of the extension primer mix. A total of 14 extension primers were used (14 SNPs being analysed) and they were consequently arranged in ascending order according to their mass. Thereafter, the primers were required to be split into 4 groups, again according to their mass, which yielded two groups of four, and two groups of 3, as detailed in the 4-step adjustment process in the table below.

4.4.6.1. iPLEX® gold extend primer mix reaction

Table 4.7 4-Step adjustment table for extension primer mix

Extend Primer Group	Final Con^{c}/primer	Volume/primer	N^{o} Primers/group	Total Vol Volume
1	7µM	8.75µL	4	35µL
2	9.3µM	11.63µL	4	46.52µL
3	11.6µM	14.58µL	3	131.22µL
4	14µM	17.5µL	3	157.5µL
Nanopure Water				379.76µL
Total Volume				**750µL**

The extension primers were stored in the PCR fridge at 5^{o}C until required. The reagents were mixed into a sterile 1,000µL eppendorf tube, and mixed using a whirlmixer® for 30 seconds. The solution was then stored back in the PCR fridge at 5^{o}C until it was required.

4.4.6.2. iPLEX® gold reaction mixture

Following the preparation of the extension primers, the iPLEX® reaction mixture was made according to the reagents and volumes outlined below.

Table 4.8 iPLEX reaction reagents and volumes

Reagent	Volume for Single Reaction	One-plex Volume	Storage Temp (^{O}C)
Nanopure water	0.755µL	400.01µL	Room
iPLEX® buffer	0.2µL	105.98µL	-20
iPLEX® termination mix	0.2µL	105.98µL	-20
Extension primer mix	0.804µL	426.06µL	-5
iPLEX® enzyme	0.04µL	21.20µL	-20
Total	**2.0µL**	**1059.34µL**	

All frozen products were allowed to thaw at room temperature, and were mixed in a sterile 5,000µL bijou tube and mixed with a whirlmixer® for 30 seconds. The extension

primer mix that was produced in the previous step (3.5.1) was taken from the -5^{o}C fridge and allowed to reach room temperature.

Into a new 96-well 'v-plate', 85µL of the iPLEX® reaction cocktail were pipetted into wells H1 – H12. Following this, an 8-tip multi-channel pipette was used to draw 10µL from row H and dispensed into each of the wells in rows A – G.

4.4.6.3. Addition of iPLEX® gold reaction mixture to DNA-containing plate

Using the automated matrix liquid handler (MLH), 2µL of the iPLEX® gold reaction cocktail was added to the 384-well DNA-containing plate as per the manufacturer's instructions.

The plate was then sealed and centrifuged for 1 minute at 2,000rpm ready for the final PCR reaction.

4.4.6.4. iPLEX® reaction

The final iPLEX® reaction was carried out on the standard PCR cycler as per the manufacturer's instructions.

4.4.6.5. iPLEX® reaction products – clean-up

The purpose of this stage is to maximise the data quality obtained via the mass spectrometry by removing excess ions and unincorporated reaction products from the IPLEX® reaction. This is done via the addition of spectroCLEAN resin and desalts the mixture via cation exchange.

SpectroCLEAN resin was applied to the left-hand side of a 6mg dispenser containing 384 wells. A Perspex plate is then used to spread the resin evenly across all 384 wells, with any excess recycled for future use. Once distributed across the dimple plate, the 384-well DNA-containing plate is inverted over the resin plate and again turned such that the resin drops into the PCR plate. This was then sealed and placed between two secure

polystyrene blocks in the Heidolph-Reax 2 rotator and allowed to rotate for 10 minutes at speed level 1 (slowest speed setting).

Once the resin-DNA mixture had been thoroughly mixed, the plate was centrifuged for 5 minutes at 3,000rpm in order to ensure the aqueous DNA mixture was the supernatant.

4.4.6.6. Dispensing of the iPLEX® reaction products into the SpectroCHIP®

Dispensing of the PCR reaction products onto the spectroCHIP® for MALDI-TOF analysis was by the automated MassARRAY® nanodispenser. This machine accurately pipettes between 15-25µL of the sample onto the matrix of the spectroCHIP®, whereby co-crystallisation occurs for the MALDI-TOF mass spectrometry stage.

The MassARRAY® nanodispenser was first prepared and conditioned as per the manufacturers protocol, and includes washing and sterilisation of the pin heads using an ultrasonic ethanol sonicator bath and pin drying. Calibrate solution was then taken from the PCR freezer (-20^{o}C) and allowed to thaw at room temperature, and 70µL pipetted into the calibrant reservoir. A spare spectroCHIP® was then used in order to ascertain the accuracy of the nanodispenser, and once this was satisfactory, a new spectroCHIP® was placed in the chip holder and the programme run to spot the chip.

4.4.7. Matrix assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS)

MALDI-TOF MS is a technique used in mass spectrometry that enables the detection and analysis of biomolecules such as proteins, sugars, DNA and oligonucleotides. The process by which the technology works is two step: first, desorption of the sample into the chip matrix is triggered by an ultraviolet laser of typical wavelength 337nm. The second step is ionisation of the ablated species, which enables acceleration by means of

an electric field. All molecules are given the same kinetic energy, and thus determination of the anylate is dependent on mass by means of its ‘time of flight’.

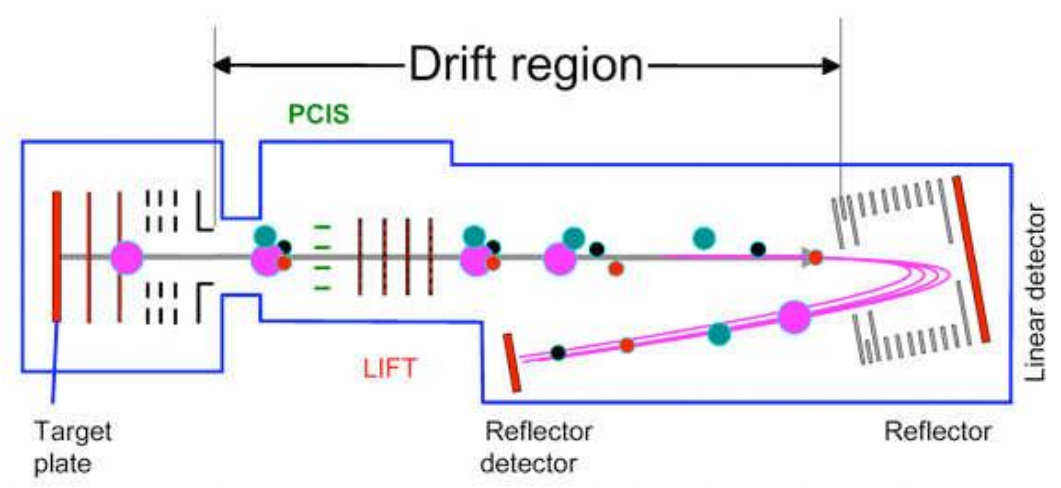

Figure 4.1 A schematic of a MALDI-TOF MS system.
Taken from. http://www.giga.ulg.ac.be/jcms/cdu_15169/maldi-tof/tof-bruker-ultraflex-ii-tof/tof-april-2005

The anylate is desorbed into the matrix and ionisation by way of a UV laser occurs. The ionised species are deflected through an electric field with their time of flight dictated by their mass.

4.4.7.1. Analysis of the SpectroCHIP®

Upon completion of sample dispensing, the chip was transferred to the MassARRAY® instrument using magnetic tongs in order to prevent contamination. The chip was placed into the mass spectrometer and the vacuum established. Once complete, the laser desorption/ionisation process was initiated as per the manufacturers protocol, and the genotypes called in real-time according to their time of flight.

4.4. Data and quality control

Blanks and repeats were incorporated into each plex as detailed previously. Blank wells containing nanopure water would be reported as having no genetic material, and the repeats were cross-referenced with the identical samples to ensure concordant results. Where this was not the case, they were repeated or excluded from the results.

Prior to statistical analyses, both clinical and genotyped data were subject to the following inclusion criteria:

4.4.1. SNP control

Only those SNPs successfully genotyped in ≥ 90% of patients were to be included. No SNPs were excluded on the basis of this criterion.

4.4.2. DNA genotype control

Only those patients in whom ≥ 90% of SNPs were genotyped were included. This resulted in the exclusion of 36 samples from the SANAD cohort and 28 samples from the Glasgow cohort.

4.4.3. Hardy-Weinberg equilibrium

In order to ensure the allele frequencies of the cohorts were representative of the general population, Hardy-Weinberg calculations were carried out on our data at the p=0.001 significance level. On this basis, one SNP (rs4147617) was excluded, resulting in 16 SNPs for analysis.

4.4.4. Matching clinical and genotype data

Both clinical and genotype data for each patient was required in order to complete the statistical analysis. When data were not present for both, the patients were removed from the analysis.

These criteria resulted in the following patients from each cohort eligible for statistical analysis: 552 SANAD patients and 442 Glasgow patients.

4.5. Statistical Methods

4.5.1. Clinical covariates and data coding

The clinical covariates included in the analysis reflect those found to be significantly associated with outcomes in the earlier SANAD GWAS analysis. Data for neurological deficit and neuroimaging was not available in the Glasgow database and were therefore excluded. The included covariates were:

- Age
- Sex
- Epilepsy type (partial, generalised, unclassifiable)
- Number of pre-treatment seizures (2, 3, 4, 5, >5)
- AED treatment (sodium valproate or gabapentin vs. neither)
- EEG findings (normal, non-specific, epileptiform, not-done)

Following collection of the experimental and clinical data, the genotypes and clinical covariates were coded as detailed in table.

Table 4.9 Data coding for statistical analysis

Variable	Coding reference
Age	Age in years at first AED treatment
Sex	0 – male; 1 – female
Epilepsy type	LRE – 1; IGE – 2; UNC – 3
N^{o} pre-treatment seizures	2 – 1; 3 – 2; 4 – 3; 5 – 4; >5 - 5
AED treatment	VPA/GBP – 0; neither – 1
EEG report	Normal – 1; non-specific – 2; epileptiform – 3; not-done - 4
Genotype	Homozygous wild – 0; heterozygous – 1; homozygous mutant – 2
Time to 12-month remission	Earliest of days to 12-month remission and days to end of follow-up (censoring)
Remission during follow-up?	No – 0 (censored observations); yes – 1 (events)

LRE: localisation-related epilepsy; IGE: idiopathic generalised epilepsy; UNC – unclassifiable epilepsy; AED: antiepileptic drug; VPA: sodium valproate; GBP: gabapentin; EEG: electroencephalogram

4.5.2. Statistical analysis

Two outcomes were investigated:

1) Remission of epilepsy: 'yes' or 'no'
2) Time to 12-month remission

Analyses of association between each SNP and the first outcome were undertaken using a binary logistic regression model. For each SNP in turn, two models were fitted: the 'baseline model', which included all of the six clinical characteristics recorded at baseline as covariates (outlined previously), and the 'genotype model', which was the same as the baseline model but also included a covariate to represent the SNP of interest. The two models were compared using the likelihood ratio test, with a significant result suggesting a statistically significant association between the SNP and outcome.

Analyses of association between each SNP and time to 12-month remission statistics were carried out using Cox proportional hazards multiple regression. Again, for each SNP in turn a baseline model including the six clinical covariates adjusted for above, and a genotype model which also included the SNP of interest, were fitted and compared using the likelihood ratio test. Kaplan-Meir (KM) plots were prepared for significant associations to illustrate the log survival function between the genotype groups. KM plots are a graphical representation of the survival function or time to event which, in this instance, is the time to 12-month remission of seizures. At any point, the graph illustrates the proportion of patients free from disease (seizures) and, therefore, a plot that tends to 0 quickly over time indicates a worse prognosis.

For stage one of the analysis, the tests of association were conducted in the SANAD cohort whilst for stage two the analyses were replicated in the Glasgow cohort. Finally, for stage three, a pooled analysis including both cohorts was undertaken. In the latter, an additional covariate representing study cohort (SANAD = 0; Glasgow = 1) was introduced to the baseline and genotype models. For each stage, the false discovery rate (FDR) was calculated for each association to account for the multiplicity of tests undertaken. A FDR<0.05 was assumed to represent statistical significance. The FDR were calculated using the 'p.adjust' function in the statistical package, 'R'.

Chapter 5

Results

Chapter 5: Results

The statistical package SPSS (version 20) was used to conduct the statistical analyses of association.

5.1. Linkage disequilibrium

Linkage disequilibrium (LD) is the non-random relationship observed between alleles at different chromosomal loci. Below are LD plots for each of the SANAD and Glasgow patient cohorts, including a plot for the combination of both datasets.

HaploView version 4.2 was used in order to create an LD plot of the SNPs used in this analysis. The numbers in each square represent the percentage probability that those SNPs are inherited together – that is, how much they are in linkage disequilibrium with one another.

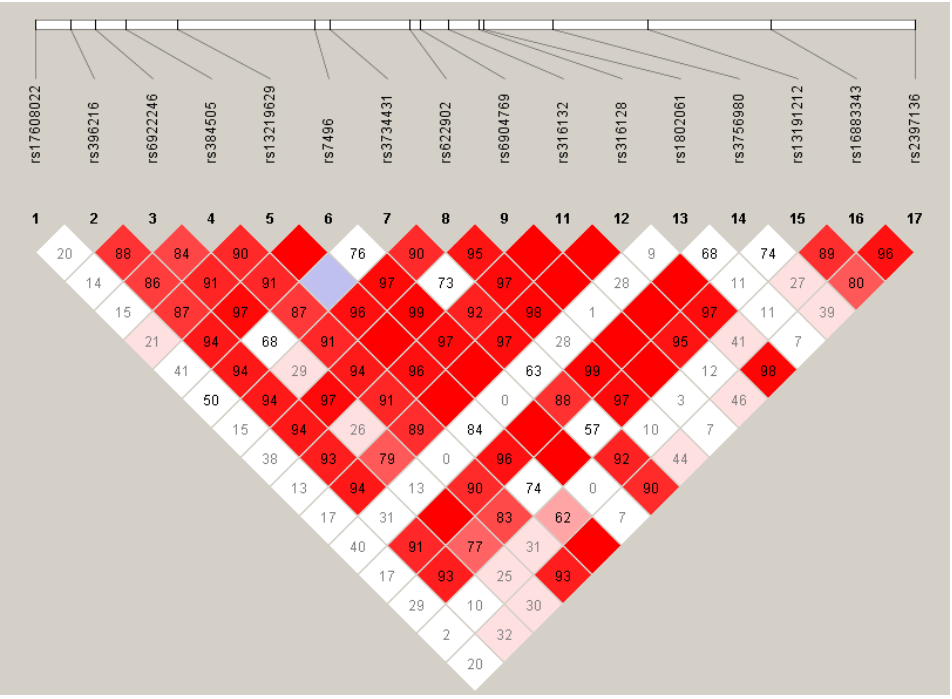

Figure 5.1. SANAD cohort LD plot

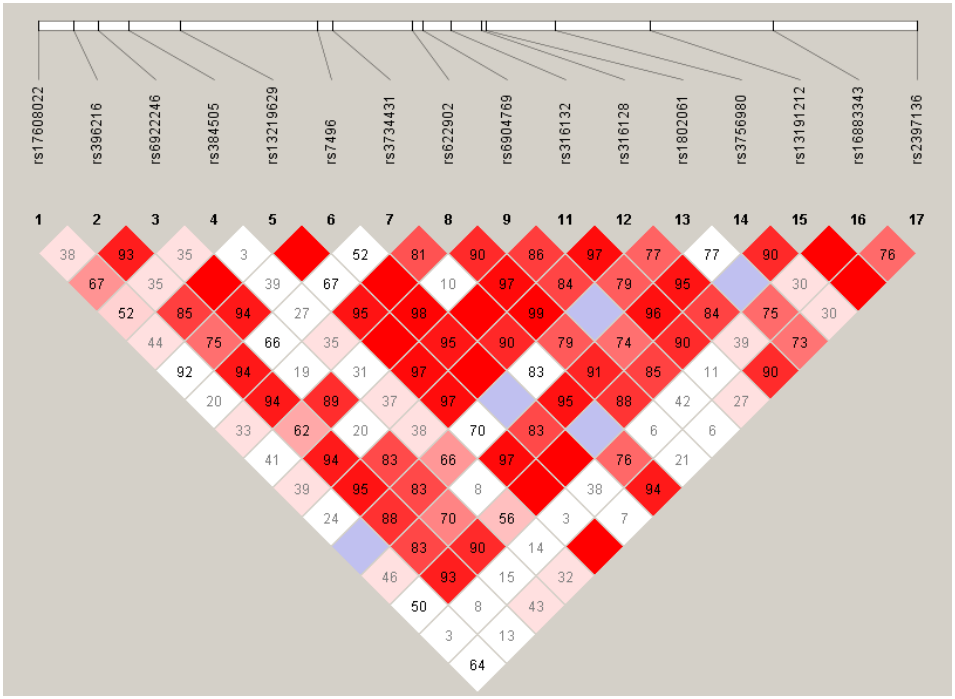

Figure 5.2. Glasgow cohort LD plot

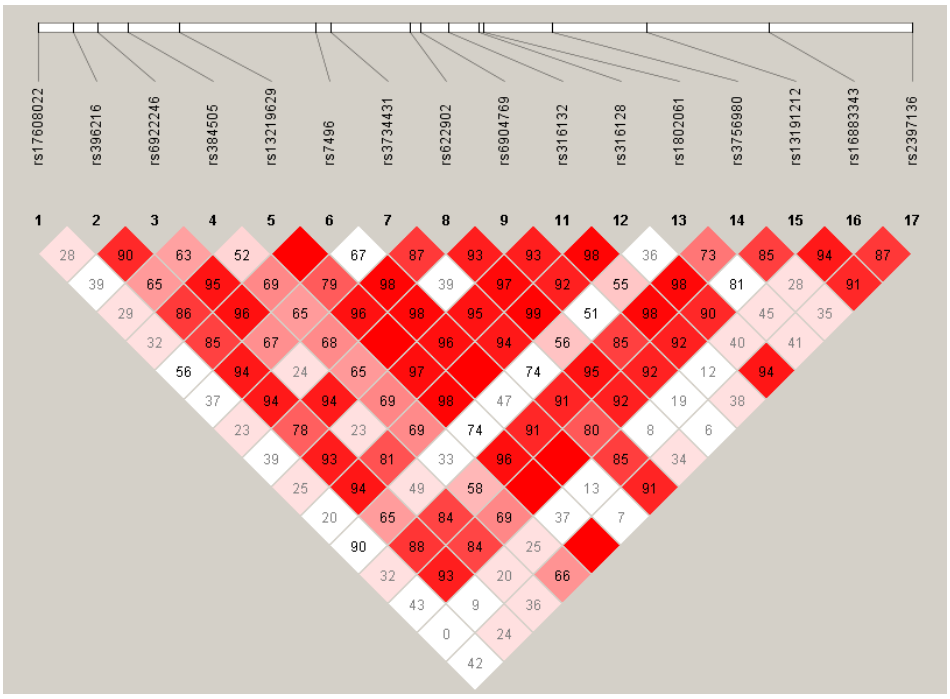

Figure 5.3. SANAD and Glasgow combined LD plot

5.2. Descriptive statistics

5.2.1. SANAD Cohort

Five-hundred and forty-one patients constituted the SANAD dataset. Of these, 287 (53.0%) were male and 254 (47.0%) were female. The range of ages amongst the patients was 18 – 99 years with a mean age of 64 years and a standard deviation of 17.8 years. For the majority of patients, their diagnosis was partial/localisation-related epilepsy with a total of 364 (67.3%) patients. Idiopathic generalised epilepsy was diagnosed in 84 (15.5%) patients, and the remaining 93 (17.2%) patients' epilepsy was unclassifiable. The number of pre-treatment seizures ranged from 2 – 10,003, with a median of 12. Sodium valproate or gabapentin was prescribed as the AED of choice in 137 (25.3%) patients, with the remaining 404 (74.7%) patients being prescribed a different AED. EEG findings were reported as normal in 225 (41.6%) patients, non-specific in 136 (25.1%) patients, epileptiform in 135 (25%) patients, and was not carried out in 45 (8.3%) of patients. Of this cohort, a total of 323 (59.7%) patients went into remission of their epilepsy during the follow-up period. The median time to 12-month remission was 538 days.

Table 5.1 Descriptive data for the SANAD dataset

Clinical Covariate		
Sex	Male	287 (53.0%)
	Female	254 (47.0%)
Age	Minimum	18 years
	Maximum	99 years
	Mean	64 years
	S.D.	17.8 years
Epilepsy type	IGE	84 (15.5%)
	LRE	364 (67.3%)
	UNC	93 (17.2%)
N^{o} pre-treatment seizures	Minimum	1
	Maximum	10,003
	Median	12
AED treatment	VPS or GBP	137 (25.3%)
	Neither	404 (74.7%)
EEG report	Normal	225 (41.6%)
	Non-specific	136 (25.1%)
	Epileptiform	135 (25.0%)
	Not-done	45 (8.3%)
Proportion achieving remission	Yes	323 (59.7%)
	No	218 (40.3%)
Median time to 12-month remission		538 days
		n = 541

S.D.: standard deviation; IGE: idiopathic generalised epilepsy; LRE: localisation-related epilepsy; UNC: unclassifiable epilepsy; AED: antiepileptic drug; VPS: sodium valproate; GBP: gabapentin; EEG: electroencephalogram

5.2.2. Glasgow Cohort

Three-hundred and ninety patients constituted the Glasgow dataset. Of these, 225 (57.7%) were male and 165 (42.3%) were female. The age of patients ranged from 13 – 81 years, with a mean of 37 years and a standard deviation of 17.3 years. Sixty-nine (17.7%) patients were diagnosed with idiopathic generalised epilepsy, 295 (75.6%) with localisation-related/partial epilepsy and the remaining 26 (6.7%) patients' epilepsy was unclassifiable. Sodium valproate or gabapentin was the prescribed AED in 93 (23.8%) patients, with the remaining 297 (76.2%) patients being prescribed an alternative agent. EEG testing was reported as normal in 143 (36.7%) patients, non-specific in 106 (27.2%) patients, epileptiform in 94 (24.1%) patients, and was not conducted in 47 (12.1%) patients. The number of pre-treatment seizures in the Glasgow cohort were already censored and therefore the raw data for this covariate was not available. A total of 278 (71.3%) patients achieved seizure remission, with a median time to 12-month remission of 550 days.

Table 5.2. Descriptive data for the Glasgow dataset

Clinical Covariate		
Sex	Male	225 (57.7%)
	Female	165 (42.3%)
Age	Minimum	13 years
	Maximum	81 years
	Mean	37 years
	S.D.	17.3 years
Epilepsy type	IGE	69 (17.7%)
	LRE	295 (75.6%)
	UNC	26 (6.7%)
AED treatment	VPS or GBP	93 (23.8%)
	Neither	297 (76.2%)
EEG report	Normal	143 (36.7%)
	Non-specific	106 (27.2%)
	Epileptiform	94 (24.1%)
	Not-done	47 (12.1%)
Proportion achieving remission	Yes	278 (71.3%)
	No	112 (28.7%)
Median time to 12-month remission		550 days
		n=390

S.D.: standard deviation; IGE: idiopathic generalised epilepsy; LRE: localisation-related epilepsy; UNC: unclassifiable epilepsy; AED: antiepileptic drug; VPS: sodium valproate; GBP: gabapentin; EEG: electroencephalogram

5.3. Binary logistic regression analysis

Table 5.3. SANAD cohort

SNP	-2 log likelihood baseline model	-2 log likelihood genotype model	p-value (likelihood test)	FDR	OR	95% CI
rs7496	662.60	656.09	0.0107	0.0679	0.583	0.384 - 0.885
rs316128	660.16	653.83	0.0119	0.0543	0.720	0.557 – 0.932
rs316132	655.62	652.22	0.0652	0.199	0.783	0.604 – 1.015
rs396216	659.67	659.61	0.806	0.878	1.034	0.788 – 1.357
rs622902	661.31	658.63	0.102	0.271	0.808	0.625 – 1.044
rs2397136	662.60	660.96	0.200	0.427	0.802	0.573 – 1.123
rs3734431	660.75	660.48	0.0633	0.776	0.886	0.564 – 1.392
rs3756980	656.89	656.69	0.655	0.776	1.075	0.785 – 1.474
rs6904769	661.79	653.16	0.00331	0.0353	0.535	0.351 – 0.816
rs6922246	657.47	645.94	0.000685	0.0109	0.562	0.401 – 0.786
rs13191212	654.98	654.16	0.365	0.625	0.816	0.526 – 1.265
rs13219629	657.02	656.22	0.371	0.625	1.196	0.806 – 1.776
rs16883343	655.62	655.40	0.639	0.776	1.073	0.802 – 1.436
rs384505	662.60	655.06	0.00603	0.0483	1.432	1.106 – 1.855
rs1802061	662.60	662.55	0.823	0.878	0.940	0.547 – 1.617
rs17608022	662.60	662.36	0.624	0.776	0.903	0.603 – 1.352

FDR: false discovery rate; OR: odds ratio; CI: confidence interval

Binary logistic regression analysis of the SANAD dataset found 3 SNPs (rs6904769, FDR = 0.0353, OR = 0.535, 95% CI 0.351 – 0.816; rs6922246, FDR = 0.0109, OR = 0.0109, 95% CI 0.401 – 0.786; and rs384505, FDR = 0.0483, OR = 1.432, 95% CI 1.106 – 1.855) to have a statistically significant association with remission of epilepsy.

Table 5.4. Glasgow cohort

SNP	-2 log likelihood baseline model	-2 log likelihood genotype model	p-value (likelihood test)	FDR	OR	95% CI
rs7496	438.86	438.61	0.617	0.859	1.140	0.678 – 1.918
rs316128	439.68	438.99	0.406	0.685	1.150	0.828 – 1.598
rs316132	430.42	428.95	0.225	0.609	1.225	0.881 – 1.704
rs396216	435.84	434.51	0.249	0.612	1.230	0.863 – 1.753
rs622902	437.63	435.16	0.116	0.609	1.299	0.935 – 1.806
rs2397136	440.04	439.86	0.671	0.882	0.910	0.590 – 1.402
rs3734431	437.25	436.30	0.330	0.684	1.360	0.723 – 2.555
rs3756980	439.31	434.87	0.0351	0.562	1.553	1.018 – 2.371
rs6904769	438.35	438.29	0.806	0.900	1.059	0.647 – 1.735
rs6922246	439.31	439.23	0.777	0.900	0.939	0.618 – 1.426
rs13191212	440.04	438.59	0.229	0.609	0.746	0.465 – 1.198
rs13219629	440.04	436.50	0.0599	0.609	1.545	0.913 – 2.614
rs16883343	427.15	427.13	0.888	0.916	0.972	0.663 – 1.424
rs384505	440.04	439.98	0.806	0.900	0.958	0.686 – 1.338
rs1802061	440.04	438.31	0.188	0.609	0.652	0.348 – 1.222
rs17608022	440.04	439.45	0.442	0.684	1.205	0.774 – 1.951

FDR: false discovery rate; OR: odds ratio; CI: confidence interval

Following adjustment for FDR, no significant associations were observed with seizure remission in the Glasgow dataset.

Both the SANAD and Glasgow cohorts were combined in order to increase statistical power. The results of the analysis are shown below.

Table 5.5. Combined cohorts

SNP	-2 log likelihood baseline model	-2 log likelihood genotype model	p-value (likelihood test)	FDR	OR	95% CI
rs7496	1137.90	1135.32	0.108	0.266	0.775	0.568 – 1.057
rs316128	1137.46	1135.15	0.129	0.279	0.860	0.707 – 1.045
rs316132	1120.70	1120.20	0.479	0.667	0.931	0.764 – 1.135
rs396216	1131.52	1130.67	0.357	0.543	1.103	0.895 – 1.360
rs622902	1136.36	1136.30	0.806	0.833	0.976	0.803 – 1.187
rs2397136	1139.92	1138.22	0.192	0.385	0.843	0.652 – 1.089
rs3734431	1135.66	1135.60	0.806	0.833	1.045	0.737 – 1.483
rs3756980	1132.18	1128.23	0.0469	0.150	1.277	1.001 – 1.628
rs6904769	1137.24	1133.23	0.0452	0.150	0.731	0.539 – 0.992
rs6922246	1134.73	1125.84	0.00287	0.0459	0.682	0.531 – 0.877
rs13191212	1130.02	1126.33	0.0547	0.159	0.736	0.539 – 1.005
rs13219629	1134.07	1129.70	0.0366	0.146	1.378	1.015 – 1.871
rs16883343	1119.25	1119.04	0.647	0.796	1.054	0.841 – 1.320
rs384505	1139.92	1135.22	0.0302	0.146	1.242	1.020 – 1.512
rs1802061	1139.92	1138.32	0.206	0.388	0.770	0.514 – 1.152
rs17608022	1139.92	1139.91	0.920	0.920	1.012	0.751 – 1.366

FDR: false discovery rate; OR: odds ratio; CI: confidence interval

By combining the SANAD and Glasgow cohorts, one association held significance (rs6922246, FDR = 0.0459, OR = 0.682; 95% CI 0.531 – 0.877).

5.4. Cox proportional hazards multiple regression analysis

Table 5.6. SANAD cohort

SNP	-2 log likelihood baseline model	-2 log likelihood genotype model	p-value (likelihood test)	FDR	HR	95% CI
rs7496	3677.07	3671.37	0.0169	0.0679	0.728	0.555 – 0.955
rs316128	3676.41	3673.09	0.0684	0.199	0.865	0.741 – 1.011
rs316132	3621.33	3619.13	0.138	0.315	0.887	0.757 – 1.040
rs396216	3661.97	3661.84	0.718	0.821	1.030	0.877 – 1.210
rs622902	3664.66	3662.40	0.133	0.315	0.888	0.760 – 1.037
rs2397136	3677.07	3675.99	0.299	0.597	0.895	0.725 – 1.105
rs3734431	3664.06	3663.63	0.512	0.776	0.914	0.696 – 1.201
rs3756980	3648.62	3648.42	0.655	0.776	1.045	0.865 – 1.262
rs6904769	3676.30	3669.19	0.00767	0.0491	0.698	0.529 – 0.921
rs6922246	3662.08	3647.27	0.000119	0.00380	0.744	0.601 – 0.922
rs13191212	3597.31	3596.76	0.458	0.733	0.900	0.679 – 1.192
rs13219629	3660.04	3659.11	0.335	0.625	1.121	0.891 – 1.410
rs16883343	3659.18	3658.96	0.639	0.776	1.042	0.877 – 1.238
rs384505	3677.07	3672.22	0.0276	0.0983	1.192	1.019 – 1.393
rs1802061	3677.07	3677.07	1	1	1.014	0.723 – 1.421
rs17608022	3677.07	3677.06	0.920	0.950	1.015	0.796 – 1.295

FDR: false discovery rate; HR: hazard ratio; CI confidence interval

Using a Cox proportional hazards multiple regression model on the SANAD dataset, 2 SNPs (rs6904769, FDR = 0.0491, HR = 0.698, 95% CI 0.529 – 0.921; rs6922246, FDR = 0.00380, HR = 0.744, 95% CI 0.601 – 0.922) were found to show significant association with time to 12-month remission. The KM curves for these SNPs are shown below.

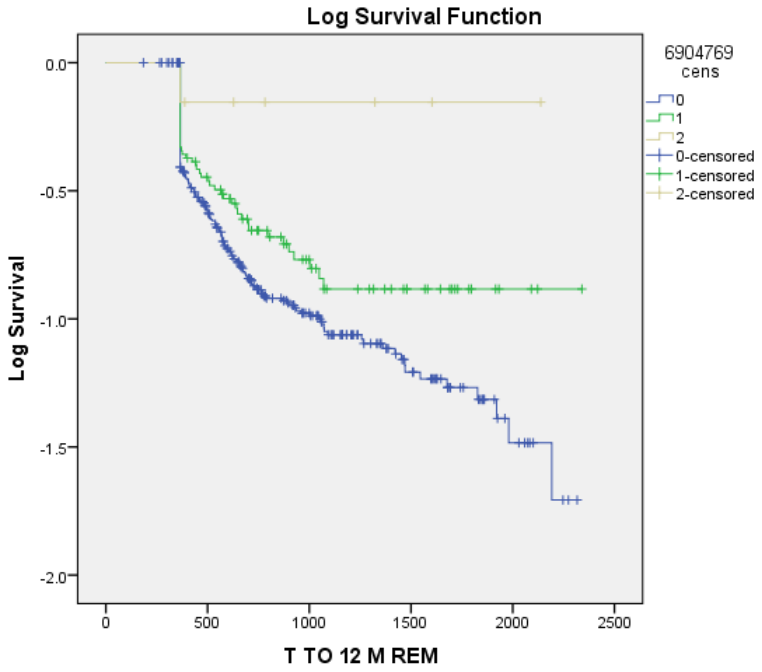

*Figure 5.4.*KM curve for rs6904769 (SANAD)

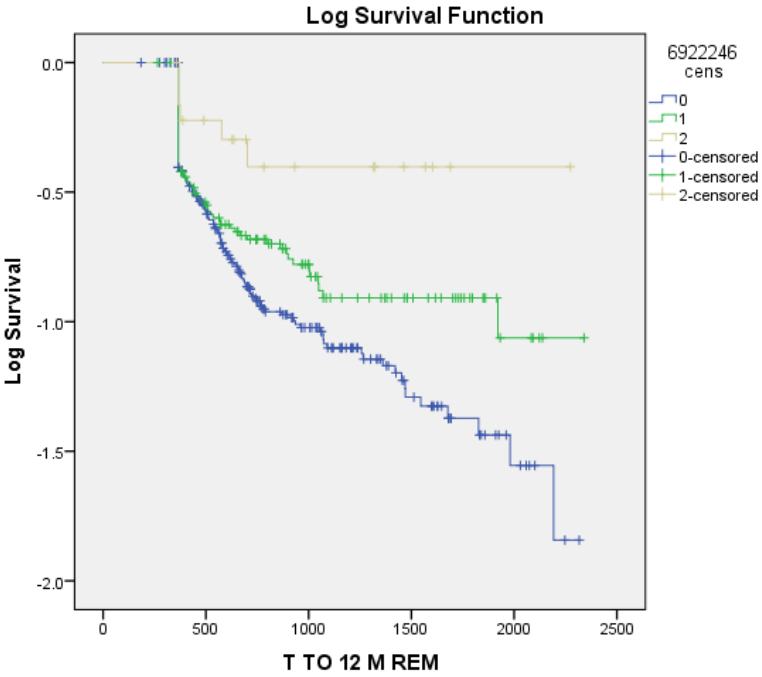

Figure 5.5 KM curve for rs6922246 (SANAD)

Figures 4.4 and *figure 4.5* are Kaplan-Meir log survival functions for SNPs rs6904769 and rs6922246 in the SANAD cohort, respectively. In both KM plots it can be seen that the patients with the mutant allele (2) are more likely to achieve seizure freedom than the heterozygous (1) and wild (0) types.

Table 5.7. Glasgow cohort

SNP	-2 log likelihood baseline model	-2 log likelihood genotype model	p-value (likelihood test)	FDR	HR	95% CI
rs7496	2934.83	2934.67	0.689	0.882	0.947	0.729 – 1.231
rs316128	2946.27	2945.73	0.462	0.685	1.067	0.897 – 1.269
rs316132	2858.79	2857.20	0.207	0.609	1.118	0.940 – 1.330
rs396216	2933.23	2931.07	0.141	0.609	1.151	0.956 – 1.386
rs622902	2933.92	2932.21	0.191	0.609	1.122	0.944 – 1.333
rs2397136	2957.76	2957.72	0.841	0.898	0.978	0.776 – 1.232
rs3734431	2943.77	2941.73	0.153	0.609	1.259	0.928 – 1.708
rs3756980	2946.24	2940.76	0.0192	0.562	1.282	1.050 – 1.565
rs6904769	2945.43	2944.63	0.371	0.685	0.892	0.692 – 1.151
rs6922246	2946.24	2945.26	0.332	0.685	0.899	0.726 – 1.113
rs13191212	2957.76	2957.19	0.450	0.685	0.897	0.674 – 1.193
rs13219629	2946.24	2943.31	0.0869	0.609	1.240	0.975 – 1.578
rs16883343	2878.84	2878.77	0.791	0.900	1.028	0.844 – 1.253
rs384505	2957.76	2957.75	0.920	0.920	0.995	0.837 – 1.181
rs1802061	2957.76	2957.19	0.450	0.685	0.873	0.609 – 1.252
rs17608022	2957.76	2957.24	0.471	0.685	1.097	0.856 – 1.404

FDR: false discovery rate; HR: hazard ratio; CI: confidence interval

Cox proportional hazards multiple regression analysis on the Glasgow dataset failed to find significance in any of the 16 SNPs tested.

Table 5.8. Combined cohorts

To further increase statistical power, the SANAD and Glasgow cohorts were once again combined. The results of this are shown below.

SNP	-2 log likelihood baseline model	-2 log likelihood genotype model	p-value (likelihood test)	FDR	HR	95% CI
rs7496	7401.04	7396.61	0.0353	0.146	0.821	0.680 – 0.991
rs316128	7413.28	7412.02	0.262	0.441	0.936	0.835 – 1.051
rs316132	7256.36	7256.20	0.689	0.817	0.977	0.870 – 1.096
rs396216	7382.69	7381.24	0.229	0.406	1.077	0.955 – 1.215
rs622902	7386.65	7386.56	0.764	0.833	0.983	0.877 – 1.102
rs2397136	7427.11	7426.19	0.337	0.540	0.927	0.794 – 1.084
rs3734431	7397.70	7397.29	0.522	0.695	1.070	0.873 – 1.312
rs3756980	7382.39	7377.02	0.0205	0.146	1.178	1.027 – 1.350
rs6904769	7412.24	7405.32	0.00852	0.0909	0.783	0.649 – 0.945
rs6922246	7385.15	7375.98	0.00246	0.0459	0.797	0.685 – 0.926
rs13191212	7338.89	7336.61	0.131	0.280	0.861	0.706 – 1.050
rs13219629	7396.63	7392.00	0.0314	0.146	1.202	1.020 – 1.417
rs16883343	7320.80	7320.25	0.458	0.667	1.050	0.923 – 1.195
rs384505	7427.11	7423.90	0.0732	0.195	1.109	0.990 – 1.243
rs1802061	7427.11	7426.74	0.543	0.695	0.928	0.727 – 1.185
rs17608022	7427.11	7427.03	0.777	0.833	1.024	0.862 – 1.216

FDR: false discovery rate; HR: hazard ratio; CI: confidence interval

Cox proportional hazard multiple regression analysis of the combined SANAD and Glasgow cohorts shows that 1 SNP (rs6922246, FDR = 0.0459, HR = 0.797, 95% CI 0.685 – 0.926) is significantly associated with time to 12-month remission. The KM curve for rs6922246 is shown below.

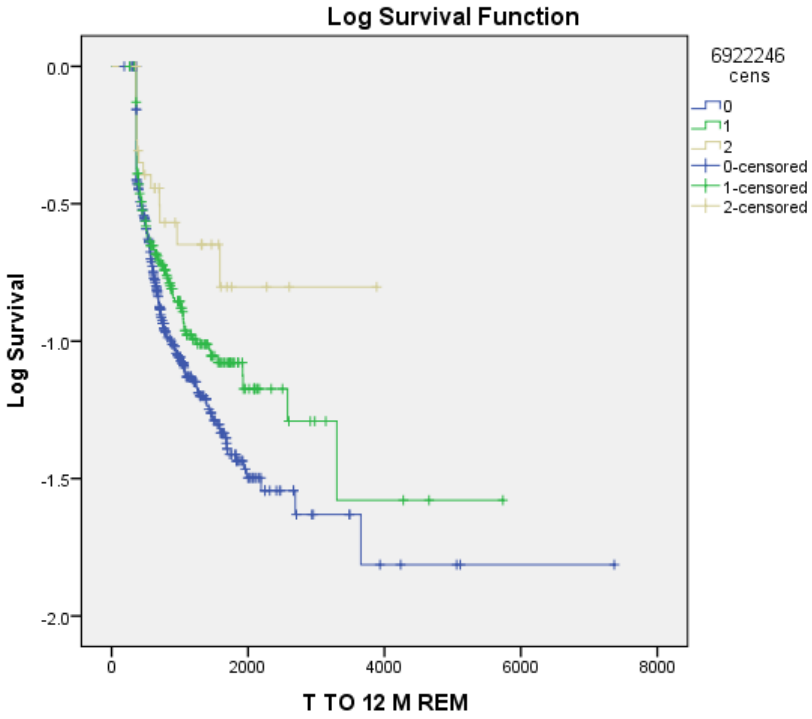

Figure 5.6. KM curve for rs6922246 (SANAD & Glasgow combined)

Figure 4.6 is a Kaplan-Meir survival plot for rs6922246 in the combined SANAD and Glasgow cohorts. Patients with the mutant allele (2) are show to have a higher degree of seizure remission when compared with the heterozygous (1) and wild (0) genotypes.

Chapter 6

Discussion

Chapter 6: Discussion

The ultimate goal of AED therapy is seizure freedom without adverse drug effects. Currently, over 30% of patients treated with several AEDs fail to become seizure-free and appear to be pharmacoresistant to the current available drugs (4). The patients that comprise the pharmacoresistant group have, by the nature of their epilepsy, been trialled on multiple AEDs with many efforts to titrate each drug to their maximum tolerated dose. Many of these efforts are attempted in vain, and commensurate with such polytherapy are often numerable side-effects that render the patient's health and quality of life impacted upon. Therefore, markers that could differentiate between AED responders, and those that are pharmacoresistant, could potentially guide clinicians in predicting which patients initiated on AED therapy would be better responders than others.

A GWAS was undertaken on the SANAD cohort of newly-diagnosed epilepsy patients whose results were meta-analysed with the AUS cohort. The primary outcome in this analysis was 1-year remission of seizures during any stage of follow-up with a minimum period of follow-up of 1 year. The results showed that SNP rs622902 (meta p-value = 1.47×10^{-7}) within the *GSTA4* gene was most strongly associated with 1-year remission of seizures (unpublished data). As a consequence, this study sought to further investigate the role of GSTA4 in epilepsy by genotyping 17 SNPs within the *GSTA4* gene that tagged for 48 SNPs representing variation across the entire gene region. Analysis was carried out on SANAD and Glasgow DNA held at the University of Liverpool, which comprised 541 and 390 patients, respectively. Genotype data was combined with known clinical data and the following clinical covariates:

- Age
- Sex
- Epilepsy type (generalised, partial or unclassified)
- Treatment (sodium valproate or gabapentin vs. neither)
- EEG (normal, non-specific, not done)

- Number of pre-treatment seizures (2, 3, 4, 5, >5)

An interesting observation from these covariates is the inclusion of gabapentin with sodium valproate. Original univariate logistic regression of the SANAD dataset illustrated that gabapentin was associated with a different outcome when compared with other antiepileptic drug – not that it was shown to be associated with a better outcome. This difference is also the reason as to why sodium valproate is included, as this was associated with a significantly different outcome compared with the other AEDs. The SANAD study showed that gabapentin was associated with a very poor outcome, whereas valproate was associated with a good outcome (10, 11). As such, both gabapentin and valproate were needed to be adjusted for in the genetic analysis.

A binary logistic regression model was fitted for categorical 'yes/no' remission and Cox proportional hazards multivariate regression analysis for time to 12-month remission.

Three SNPs were significantly associated with remission (binary outcome, yes/no) during follow-up in the SANAD cohort: rs6904769 (FDR = 0.0353), rs6922246 (FDR = 0.0109) and rs384505 (FDR = 0.0483). rs6904769 is an intronic C>T SNP found in the intra-exonic gene region and is not found in LD with any known functional SNP. Its mechanism in the role of seizure remission cannot, therefore, be deduced. Both rs6922246 and rs384505 are downstream of the *GSTA4* gene in the 3' region and are G>T and G>A SNPs, respectively. Similarly, neither of these SNPs is in LD with any known functional variant. Whilst no mechanisms for these polymorphisms can be deduced, intronic and downstream 3' SNPs may alter transcription factor binding and/or result in the reduced stability of mRNA respectively, with the possible subsequent alteration of the gene product. Altered GSTA4 protein may have several consequences, which include altered AED metabolism, or a role in the altered metabolism of the toxic lipid peroxidation product, 4-HNE. Although all three SNPs were significant in the SANAD cohort, replication in the Glasgow dataset failed. Combination of the two cohorts for

greater statistical power led to rs6922246 (FDR = 0.0459) being the only SNP associated with this phenotype.

Two SNPs were significantly associated with time to 12-month remission: rs6904769 (FDR = 0.0491) and rs6922246 (FDR = 0.00380). These associations were again seen in the SANAD dataset only, with failure of replication in the Glasgow cohort. In combining the two populations for greater statistical power, only one of the statistical associations held – rs6922246 (FDR = 0.0459).

The fact that both rs6922246 and rs384505 are found in the downstream 3' region of the *GSTA4* gene may be of significance and would point towards a potential role of these variants in mRNA stability or susceptibility to degradation by micro-RNAs. How this then relates to the likelihood of achieving a remission from seizures is unclear.

It is interesting to note that the association between remission of seizures following initiation of AED therapy and rs384505 was also seen in the GWAS conducted on the AUS cohort, and was the rationale for this SNP being included into our SNP selection. Whilst the rs384505 SNP did not show a significant association after combining the SANAD and Glasgow cohorts, replication was seen in the SANAD dataset. It is therefore reasonable to suggest that rs384505 may have a role in the remission of epilepsy, though further analysis is required.

The original GWAS meta-analysis showed the most significant association with 1-year remission phenotype at the rs622902 loci. In our findings, this SNP did not show any statistical association in either outcome, even though genotyping was carried out on a subset of the same SANAD cohort. However, in the GWAS, significance at the loci of rs622902 may have simply represented significant polymorphic variation at the *GSTA4* region, which we have replicated, albeit in different SNPs. It is therefore reasonable to suggest that the findings in this study are more robust, given the denser genotyping conducted at this specific genomic region only.

In all of the statistical analyses conducted in this study, it appears that the greatest association between seizure remission is seen with rs6922246. Significant associations in the SANAD cohort were seen in both statistical models, and whilst replication was not seen in the Glasgow dataset, the associations held for the combined cohorts. This suggests that there may be a significant role for rs6922246 in newly-diagnosed epilepsy, and it is therefore a strong candidate for further validation.

There are several hypotheses as to why the associations were not replicated in the Glasgow cohort. As stated, the Glasgow patients were representative of a discrete geographical area whose GSTA4 SNP variation may be different from the more extensive nation-wide population that was the SANAD cohort. A further issue exists in the diagnostic process of epilepsy; unlike most diseases that utilise GWAS, epilepsy is a diagnosis based upon clinical acumen with no definitive biochemical test. Therefore, in order to truly compare both cohorts, one must consider the possibility of variation in the clinical diagnosis of the condition between these groups, which could have significant effects on the results of a genome-wide association study. Combined, these limitations could substantially influence the results obtained in a GWAS analysis.

SNP rs4147616 is a missense SNP and was, therefore, one of the force inclusion SNPs in the original list, though with MAF = 0.0069. However, analysis of the genotype data showed that there was a technical fault with the primers, as all reported genotypes were identical – this also resulted in the failure of Hardy-Weinberg equilibrium. Give missense SNPs are thought to be the most likely polymorphisms to result in a functional variant, it is disappointing that the genotype data was not available for our outcome analysis.

The neighbouring gene upstream of *GSTA4* is the intestinal cell kinase (*ICK*) gene. ICK is a protein found in the crypts of the intestinal villi and is responsible for cellular proliferation and differentiation (104). Thus, polymorphisms within *ICK* may result in the altered expression of channels in the intestine that are responsible for drug transport.

Interestingly, gabapentin is absorbed solely form the proximal small intestine and is excreted by the kidneys unaltered (105). With *GSTA4* and ICK being in close genomic proximity, it may be that SNPs in these genes are in linkage disequilibrium – that is, the SNPs flagged in the GWAS may represent association with *ICK* instead of *GSTA4*. Therefore, it is not unreasonable to suspect that polymorphic variation in *ICK* may alter the transport of gabapentin and afford some explanation to the results obtained in the earlier analysis.

It is thought that genetic variability between individuals plays at least a part in determining whether patients are drug responders, or whether they are pharmacoresistant to therapy (98). GSTA4 is a ubiquitous enzyme that belongs to the glutathione-conjugating super-family of enzymes. These enzymes, and in particular GSTA4, have been shown to be responsible for the metabolism of toxic by-products that are produced by lipid peroxidation reactions, namely the metabolite 4-HNE (92). In abnormal tissues that are subject to high cellular stress, 4-HNE is found in greater levels, and may be used as a biomarker for cellular stress (92, 93). This may be the case in epileptogenic tissue and may explain why GSTA4 has a role in epilepsy. Further, AEDs may lead to an increase 4-HNE concentration that, in abnormal epileptogenic tissue, may exacerbate seizures and result in pharmacoresistance. Upregulation of GSTA4 in tissues containing high 4-HNE concentrations may contribute to increased local metabolism of AEDs and thus compound pharmacoresistance in these patients.

Unfortunately, there are no published studies that have explored the link between GSTA4 and epilepsy. Associations between epilepsy and/or AEDs and other classes of GST have been shown (81, 95-97) although replication of these data and analysis in other independent cohorts is required. The association between these SNPs and remission from seizures is, therefore, novel, and necessitates further exploration in order to determine their functional significance and whether this association holds in the wider epilepsy population.

6.1. Limitations

Clinical data obtained from the SANAD cohort that was used in this study was not originally intended for the use in the analyses of pharmacogenetic studies. Clinical report forms that were used to record clinical details were therefore not ideally designed for this purpose. Nevertheless, data was collected prospectively and is, therefore, assumed to be of high quality and free from bias. Data that was obtained from the Glasgow cohort was retrospectively collected from the hospital records. Given that no proforma was used, any details required could be obtained from the comprehensive patient notes, as outlined previously.

It is not known which baseline clinical characteristics truly affect the outcome of the phenotypes studied, and therefore which should be adjusted for as clinical covariates in the statistical analyses. In this study, only those characteristics that were shown to be statistically significant in the original SANAD dataset were included, which may not have been appropriate for the Glasgow cohort and indeed may not be truly representative of the population. Neurological deficit was one of the covariates included in the original SANAD GWAS analysis, and indeed had one of the most significant p-values ($p = 0.001$). Data regarding neurological deficit was, however, not available for the Glasgow cohort and could therefore not be included in this study. It is reasonable to believe that patients in whom neurological deficit is found may have more severe epilepsy than those without, and are therefore more likely to be pharmacoresistant to AEDs.

It is known that certain AEDs are superior to others in controlling specific types of seizure/epilepsy from trials such as SANAD. In this study, AED therapy was grouped into sodium valproate or gabapentin vs. any other and were not analysed individually. Stratified analysis of each AED needs to be carried out in order to better appreciate the role of each medication in our outcome measure. The choice of AED may also have compounded the results, as many hospitals/epileptologists have preferential medications

they use that may vary from another epilepsy centre. This would be less influential on the SANAD cohort, though it may influence the results obtained from the Glasgow patients.

Whilst the patients used in these studies represent the largest known genetic database of patients with newly-diagnosed epilepsy, the sample sizes may be insufficient to draw population-wide conclusions into these polymorphisms. The sample size may also be restricted further by the notorious difficulty in making an accurate diagnosis of epilepsy as described earlier – it is certainly not unreasonable to assume that a minority of individuals have, at least in part, an incorrect diagnosis, or have been initiated on a medication that is not entirely appropriate for their epilepsy type. Given these factors are clinician-dependant, they may be more pronounced in the Glasgow cohort and partly explain the lack of association seen in this patient group.

Given little is known about which characteristics do affect our outcome of interest, it is difficult to know which should be recorded and included in any future research. Further testing of covariates in other studies, or meta-analyses of published works regarding which baseline characteristics show significance in epilepsy outcome, could elaborate on this issue and aid in future association studies.

6.2. Summary of results

Sixteen SNPs (one discarded owing to probable technical issues/failure of Hardy-Weinberg equilibrium) that tagged for variation across the *GSTA4* gene region were tested for associations between the remission of seizures (yes/no) and the time to 12-month remission of seizures after adjusting for the six clinical covariate outlined above. Remission of epilepsy was statistically associated with SNPs rs6904769 and rs6922246 in the SANAD cohort. Of these, rs6922246 was the only SNP whose association held after combining the SANAD and Glasgow datasets for greater statistical power.

Time to 12-month remission of seizures was significantly associated with SNPs rs6904769, rs6922246 and rs384505 in the SANAD cohort. In combining the two datasets, this association held for rs6922246 only. It is therefore reasonable to conclude that there may be a link between remission of epilepsy, time to 12-month remission of epilepsy, and rs6922246 in particular, though further work to explore this role is required.

6.3. Study Impact

This study sought to determine the role of GSTA4 in newly-diagnosed epilepsy patients initiated on AED therapy. Via the selection of tagging SNPs representing variation across the *GSTA4* genomic region, and the genotype analysis of these SNPs by TaqMan q-PCR and Sequenom MALDI-TOF MS analysis, it has been shown that rs6922246 is associated with both remission of epilepsy and time to 12-month remission of epilepsy in our patient cohorts. Further, two other SNPs showed association; rs6904769 and rs384505 in the SANAD cohort for both outcomes of interest. Validation of these results is an independent cohort is required to explore this link further.

6.4. Future work

Since the NCBI was consulted in order to produce our list of SNPs within the *GSTA4* gene, there has been publication of a missense SNP with MAF = 0.0011 (rs45551133). Whilst the MAF of this SNP is below the original cut-off value of 5%, this SNP is of potential functional significance and therefore analysis for any association and our outcomes of interest would be a possible avenue for future work.

Sequencing of the *GSTA4* gene region in patients with epilepsy and healthy control subjects could potentially show novel variants between the groups with subsequent analysis regarding their function. This type of genetic analysis has been shown to be

successful in several clinical studies, and was used by Lupski et al (106) to identify polymorphisms in a small cohort diagnosed with the neuro-genetic disorder, Charcot-Marie-Tooth. The participants in the study were routinely tested for common variants known to be associated with Charcot-marrie-Tooth, the results of which were negative. Following further gene sequencing as outlined, novel SNPs were discovered. Similarly, Ley et al (107) used a sequencing method to discover 8 new genetic variants in a patient with acute myeloid leukaemia that were not present in matched, health control tissue.

Functional studies investigating how GSTA4 relates to epilepsy and AEDs could be another research area of interest. rs6922246 is a SNP located within the downstream 3' region of the gene, having possible effect on mRNA stability and protein transcription. In order to investigate this, methods that involve the expression of the mutant and wild GSTA4 protein using *E.coli* plasmids could be utilised. In the initial phase, the level of GSTA4 protein could be compared such that any mRNA instability in the mutant type would result in a decrease of protein expression. *E.coli* plasmids containing transfected GSTA4 protein are available commercially via LGC standards (http://www.lgcstandards-atcc.org/LGCAdvancedCatalogueSearch/ProductDescription/tabid/1068/Default.aspx?ATCCNum=MGC-9318&Template=mgcHumanClones).

Given the significant role of GSTA4 in the metabolism of 4-HNE and subsequent importance in the mechanisms of cellular stress (92), its upregulation in epileptogenic tissue and consequent increased metabolism of AEDs is also another area of potential interest. In the first instance, non-mutant GSTA4 protein could be used in order to assess the metabolising capabilities of certain AEDs in order to ascertain to what extent GSTA4 may metabolise these xenobiotics. If successful, studies could then be conducted in order to investigate how the rs6922246 SNP variant affects metabolism.

Finally, and with regard to the SNP findings of this study, our results require validation in a large, independent cohort that should ideally be prospective in nature in order to evade

confounding factors and bias, and be designed with a pharmacogenetic ideation in mind. It would be useful to have further information, as outlined in section 5.1, on those clinical covariates known to influence seizure outcome though this information is not yet available. Validation in an independent cohort is, for the time, difficult, as both the SANAD and Glasgow cohorts are the largest known datasets of newly-diagnosed epilepsy patients for which access to DNA is available, and thus replication is, for the moment, not possible. The SANAD-II clinical trial is in the phases of development, and later pharmacogenetic analysis in this cohort may be possible in order to attempt to replicate our findings, and would provide another large population in which to conduct further pharmacogenetic analysis into newly-diagnosed epilepsy.

6.5. Conclusion

The pharmacoresistant phenotype is seen in approximately 30% of the epilepsy population and represents a significant global health issue. Despite the advent of newer antiepileptic agents, there has been no significant move forward in the clinical management of these patients (3, 4). Genetic variation between individuals is thought to contribute to this phenotype, and it was the purpose of this study to investigate the role of SNPs in the *GSTA4* gene with outcome measurements being remission of seizures (yes/no) and time to 12-month remission.

The remission of seizures was associated with rs6904769 and rs6922246 in the SANAD cohort, although after combining both cohorts for statistical power association was only seen in rs6922246.

Time to 12-month remission was associated with rs6904769, rs6922246 and rs384505 in the SANAD cohort. After combining both datasets, only rs6922246 was significantly associated with this outcome.

In conclusion, we have found that, particularly, rs6922246 is associated with both of our outcome measures. Validation of our findings in an independent cohort is necessary to confirm the role of this SNP variant in the *GSTA4* gene, though the limitations of this have been described. Future research into GSTA4 and its genetic variants are necessitated to further our knowledge of its role in epilepsy, with the ultimate aim being the development of a gene marker that could be used in clinical practise to predict the response patients may have to AED therapy.

References

1. Shorvon SD. The causes of epilepsy: changing concepts of etiology of epilepsy over the past 150 years. Epilepsia. 2011;52(6):1033-44. Epub 2011/04/13.

2. Shorvon S. Handbook of Epilepsy Treatment. Second ed: Blackwell; 2005.

3. Wahab A, Albus K, Gabriel S, Heinemann U. In search of models of pharmacoresistant epilepsy. Epilepsia. 2010;51 Suppl 3:154-9. Epub 2010/07/22.

4. Pati S, Alexopoulos AV. Pharmacoresistant epilepsy: from pathogenesis to current and emerging therapies. Cleve Clin J Med. 2010;77(7):457-67. Epub 2010/07/06.

5. Ngugi AK, Kariuki SM, Bottomley C, Kleinschmidt I, Sander JW, Newton CR. Incidence of epilepsy: a systematic review and meta-analysis. Neurology. 2011;77(10):1005-12. Epub 2011/09/07.

6. Duncan JS, Sander JW, Sisodiya SM, Walker MC. Adult epilepsy. Lancet. 2006;367(9516):1087-100. Epub 2006/04/04.

7. Excellence NIfHaC. The epilepsies: the diagnosis and management of the epilepsies in adults and children in primary and secondary care 2012 [cited 2012 14/05/2012]; Available from: http://publications.nice.org.uk/the-epilepsies-the-diagnosis-and-management-of-the-epilepsies-in-adults-and-children-in-primary-and-cg137.

8. Berg AT, Berkovic SF, Brodie MJ, Buchhalter J, Cross JH, van Emde Boas W, et al. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009. Epilepsia. 2010;51(4):676-85. Epub 2010/03/04.

9. Arzimanoglou A, Ben-Menachem E, Cramer J, Glauser T, Seeruthun R, Harrison M. The evolution of antiepileptic drug development and regulation. Epileptic Disord. 2010;12(1):3-15. Epub 2010/03/10.

10. Marson AG, Al-Kharusi AM, Alwaidh M, Appleton R, Baker GA, Chadwick DW, et al. The SANAD study of effectiveness of valproate, lamotrigine, or topiramate for

generalised and unclassifiable epilepsy: an unblinded randomised controlled trial. Lancet. 2007;369(9566):1016-26. Epub 2007/03/27.

11. Marson AG, Al-Kharusi AM, Alwaidh M, Appleton R, Baker GA, Chadwick DW, et al. The SANAD study of effectiveness of carbamazepine, gabapentin, lamotrigine, oxcarbazepine, or topiramate for treatment of partial epilepsy: an unblinded randomised controlled trial. Lancet. 2007;369(9566):1000-15. Epub 2007/03/27.

12. Hardy J, Singleton A. Genomewide association studies and human disease. The New England journal of medicine. 2009;360(17):1759-68. Epub 2009/04/17.

13. Maffie J, Rudy B. Weighing the evidence for a ternary protein complex mediating A-type K+ currents in neurons. J Physiol. 2008;586(Pt 23):5609-23. Epub 2008/10/11.

14. Hopkins FG, Dixon M. On Glutathione. II. A Thermostable Oxidation-Reduction System. The Journal of Biological Chemistry. 1922;54:527-63.

15. THOMAS JA. OXIDATIVE STRESS: INCLUDING GLUTATHIONE, A PEPTIDE FOR CELLULAR DEFENSE AGAINST OXIDATIVE STRESS. 1999 [07/09/2011]; Available from: http://www.bb.iastate.edu/~jat/glutchp.html.

16. Masella R, Mazza G. Glutathione and Sulfur Amino Acids in Human Health and Disease: Wiley; 2009.

17. Schafer FQ, Buettner GR. Redox Environment of the Cell as Viewed Through The Redox State of the Glutathione Disulfide/Glutathione Couple. Free Radical Biology and Medicine. 2001;30(11):1191-212.

18. Schafer FQ, Buettner GR. Redox environment of the cell as viewed through the redox state of the glutathione disulfide/glutathione couple. Free Radical Biology & Medicine. 2001;30(11):1191-212.

19. Oakley A. Glutathione transferases: a structural perspective. Drug Metabolism Reviews. 2011;43(2):138-51.

20. Vina J, Hems R, Krebs HA. Maintenance of glutathione content in isolated hepatocyes. Biochemical Journal. 1978;170(3):627-30.

21. Gibson GG, Skett P. Introduction to Drug Metabolism. Third ed: Nelson Thornes Ltd; 2001.

22. Hayes JD, Flanagan JU, Jowsey IR. Glutathione transferases. Annual review of pharmacology and toxicology. 2005;45:51-88.

23. Rushmore TH, Pickett CB. Glutathione S-Transferases, Structure, Regulation, and Therapeutic Implications. The Journal of Biological Chemistry. 1993;268(16):11475-8.

24. Cazenave LA, Moscow JA, Myers CE, Cowan KH. Glutathione S-transferase and drug resistance. Cancer treatment and research. 1989;48:171-87. Epub 1989/01/01.

25. Combes B, Stakelum GS. A liver enzyme that conjugates sulfobromophthalein sodium with glutathione. The journal of clinical investigation. 1961;40(6):981-8.

26. Booth J, Boyland E, Sims P. An enzyme from rat liver catalysing conjugations with glutathione. Biochemistry journal. 1961;79(3):516-24.

27. Mannervik B, Board PG, Hayes JD, Listowsky I, Pearson WR. Nomenclature for mammalian soluble glutathione transferases. Methods in Enzymology. 2005;401:1-8.

28. Clark AG, Smith JN, Speir TW. Cross specificity in some vertebrate and insect glutathione-transferases with methyl parathion (dimethyl *p*-nitrophenyl phosphorothionate), 1-chloro-2,4-dinitrobenzene and *S*-crotonyl-*N*-acetylcysteamine as substrates. Biochemistry journal. 1973;135:385-92.

29. Mannervik B, Awasthi YC, Board PG, Hayes JD, Ilio CD, Ketterer B, et al. Nomenclature for human glutathione transferases. Biochemical Journal. 1992;282:305-8.

30. Morel F, Aninat C. The glutathione transferase kappa family. Drug Metabolism Reviews. 2011;43(2):281-91.

31. Reinemer P, Dirr HW, Ladenstein R, Schaffer J, Gallay O, Huber R. The three-dimentional structre of class pi glutathione S-transferase in complex with glutathione sulfonate at 2.3 A resolution. The EMBO Journal. 1991;10(8):1997-2005.

32. Dourado DFAR, Fernandes PA, Ramos MJ. Glutathione transferase classes alpha, pi, and mu: GSH activation mechanism. The journal of physical chemistry B. 2010;114:12972-80.

33. Townsend DM, Tew KD. Cancer drugs, genetic variation and the glutathione S-transferase gene family. American Journal of Pharmacogenomics. 2003;3(3):157-72.

34. Armstrong RN. Structure, catalytic mechanism, and evolution of the glutathione transferases. Chemical research in toxicology. 1997;10:2-18.

35. Hayes JD, Pulford DJ. The glutathione S-transferase supergene family: Regulation of GST and the contribution of their isoenzymes to cancer chemoprotection and drug resistance. Critical reviews in biochemistry and molecular biology. 1995;30(6):445-600.

36. Jfdwolff. Eicosanoid synthesis. 2007 [20/09/2011]; Available from: http://en.wikipedia.org/wiki/File:Eicosanoid_synthesis.svg.

37. Board PG, Coggan M, Wilce MCJ, Parker MW. Evidence for an essential serine residue in the active site of the Theta class glutathione transferases. Biochemical Journal. 1995;311:247-50.

38. Gillham B. The reaction of aralkyl sulphate esters with glutathione catalysed by rat liver preperations. Biochemical journal. 1971;121:667-72.

39. Hiratsuka A, Okada T, Nishiyama T, Fujikawa M, Ogura K, Okuda H, et al. Novel theta class glutathione S-transferase yrs-yrs' and yrs'-yrs' in rat liver cytosol: Their potent activity toward 5-sulfoxymethylchrysene, a reactive metabolite of the carcinogen 5-hydroxymethylchrysene. Biochemical and biophysical research communications. 1994;202(1):278-84.

40. Board PG, Coggan M, Chelvanayagam G, Easteal S, Jermin LS, Schulte GK, et al. Identification, characterization, and crystal structure of the omega class glutathione transferases. The journal of biological chemistry. 2000;275(32):24798-806.

41. Cho S-G, Hee-Lee Y, Park H-S, Ryoo K, Kang KW, Park J, et al. Glutathione S-transferase mu modulates the stress-activated signals by suppressing apoptosis signal-regulating kinase 1. The journal of biological chemistry. 2000;276(16):12749-55.

42. Adler V, Yin Z, Fuchs SY, Benezra M, Rosario L, Tew KD, et al. Regulation of JNK signaling by GSTP. The EMBO Journal. 1999;18(5):1321-34.

43. Kyriakis JM, Banerjee P, Nikolakaki E, Dai T, Rubie EA, Ahmad MF, et al. The stress-activated protein kinase subfamily of c-Jun kinases. Nature. 1994;369:156-60.

44. Bruns CM, Hubatsch I, Ridderstrom M, Mannervik B, Tainer JA. Human glutathione transferase A4-4 crystal structures and mutagenesis reveal the basis of high catalytic efficiency with toxic lipid peroxidation products. Journal of molecular biology. 1999;288:427-39.

45. Hayes JD, McLellan LI. Glutathione and glutathione-dependent enzymes represent a co-ordinately regulated defense against oxidative stress. Free radical research. 1999;31(4):273-300.

46. Harris JM, Meyer DJ, Coles B, Ketterer B. A novel glutathione transferase (13-13) isolated from the matrix of rat liver mitochondria having structural similarity to class Theta enzymes. Biochemical Journal. 1991;278:137-41.

47. Thomson RE, Bigley AL, Foster JR, Jowsey IR, Elcombe CR, Orton TC, et al. Tissue-specific expression and subcellular distribution of murine glutathione S-transferase class kappa. Journal of Histochemistry & Cytochemistry. 2004;52(5):653-62.

48. Raza H, Robin M-A, Fang J-K, Avadhani NG. Multiple isoforms of mitochondrial glutathione S-transferases and their differential induction under oxadative stress. Biochemical Journal. 2002;366:45-55.

49. Liu M, Zhou L, Xu A, Lam KSL, Wetzel MD, Xiang R, et al. A disulfide-bond A oxidoreductase-like protein (DsbA-L) regulates adiponectin multimerization. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(47):18302-7.

50. Oh DK, Ciaraldi T, Henry RR. Adiponectin in health and disease. Diabetes, Obesity and Metabolism. 2007;9:282-9.

51. Zhao L, Liu M, Zhang J, Chen H, Dong LQ, Liu F. DsbA-L alleviates endoscopic reticulum stress-induced adiponectin downregulation. Diabetes. 2010;59(11):2809-16.

52. Jakobsson P-J, Morgenstern R, Mancini J, Ford-Hutchinson A, Persson B. Common structural features of MAPEG - A widespread superfamily of membrane associated proteins with highly divergent functions in eicosanoid and glutathione metabolism. Protein Science. 1998;8:689-92.

53. Jakobsson P-J, Thoren S, Morgenstern R, Samuelsson B. Identification of human prostaglandin E synthase: a microsomal, glutathione-dependent, inducible enzyme, constituting a potential novel drug target. Proceedings of the National Academy of Sciences of the United States of America. 1999;96:7220-5.

54. Morel F, Rauch C, Coles B, Ferrec EL, Guillouzo A. The human glutathione transferase alpha locus: genomic organization of the gene cluster and functional characterisation of the genetic polymorphism in the hGSTA1 promoter. Pharmacogenetics. 2002;12:277-86.

55. Board PG, Webb G. Isolation of a cDNA clone and localization of human glutathione S-transferase 2 genes to chromosome band 6p12. Proceedings of the National Academy of Sciences of the United States of America. 1987;84:2377-81.

56. Islam MQ, Platz A, Szpirer J, Szpirer C, Levan G, Mannervik B. Chromosomal localization of human glutathione transferase genes of classes alpha, mu and pi. Human Genetics. 1989;82:338-42.

57. Desmots F, Rissel M, Loyer P, Turlin B, Guillouzo A. Immunohistological analysis of glutathione transferase A4 distribution in several human tissues using a specific polyclonal antibody. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society. 2001;49(12):1573-80. Epub 2001/11/29.

58. Ng L, Nichols K, O'Rourke K, Maslen A, Kirby GM. Repression of human GSTA1 by interleukin-1beta is mediated by variant hepatic nuclear factor-1C. Molecular pharmacology. 2007;71(1):201-8. Epub 2006/10/06.

59. Bredschneider M, Klein K, Murdter T, Marx C, Eichelbaum M, Nussler A, et al. Genetic polymorphisms of glutathione S-transferase A1, the major glutathione S-transferase in human liver: consequences for enzyme expression and busulfan conjugation. Clinical Pharmacology and Therapeutics. 2002;71(6):479-87.

60. Ilic Z, Crawford D, Vakharia D, Egner PA, Sell S. Glutathione-S-transferase A3 knockout mice are sensitive to acute cytotoxic and genotoxic effects of aflatoxin B1. Toxicology and applied pharmacology. 2010;242(3):241-6. Epub 2009/10/24.

61. Johansson A-S, Mannervik B. Human glutathione transferase A3-3, a highly efficient catalyst of double-bond isomerization in the biosynthetic pathway of steroid hormones. The journal of biological chemistry. 2001;276(35):33061-5.

62. Board PG. Identification of cDNAs encoding two human alpha class glutathione transferases (GSTA3 and GSTA4) and the heterologous expression of GSTA4-4. Biochemical journal. 1998;330:827-31.

63. Desmots F, Rissel M, Loyer P, Turlin B, Guillouzo A. Immunohistological analysis of glutathione transferase A4 distribution in several human tissues using a specific polyclonal antibody. Journal of histochemistry & cytochemistry. 2001;49(12):1573-9.

64. Hayes JD, Pulford DJ, Ellis EM, McLeod R, James RFL, Seidegard J, et al. Regulation of rat glutathione S-transferase A5 by cancer chemopreventive agents: Mechanisms of inducible resistance to aflatoxin B_1. Chemico-Biological Interactions. 1998;111-112:51-67.

65. Pulford DJ, Hayes JD. Characterization of the rat glutathione S-transferase Yc_2 subunit gene, *GSTA5*: identification of a putative antioxidant-responsive element in the 5'-flanking region of rat *GSTA5* that may mediate chemoprotection against aflatoxin B_1. Biochemical Journal. 1996;318:75-84.

66. Pearson WR, Vorachek WR, Xu S-j, Berger R, Hart I, Vannais D, et al. Identification of class-mu glutathione transferase genes GSTM1-GSTM5 on human chromosome 1p13. The American journal of human genetics. 1993;53:220-33.

67. Board PG. Biochemical genetics of glutathione-S-transferase in man. American journal of human genetics. 1981;33(1):36-43. Epub 1981/01/01.

68. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet. 2003;33(2):177-82. Epub 2003/01/14.

69. Huang RS, Chen P, Wisel S, Duan S, Zhang W, Cook EH, et al. Population-specific GSTM1 copy number variation. Human molecular genetics. 2009;18(2):366-72. Epub 2008/10/25.

70. Laisney V, Cong NV, Gross MS, Frezal J. Human genes for glutathione s-transferases. Human Genetics. 1984;68:221-7.

71. Liloglou T, Walters M, Maloney P, Youngson J, Field JK. A T2517C polymorphism in the *GSTM4* gene is associated with risk of developing lung cancer. Lung Cancer. 2002;37(2):143-6.

72. Takahashi Y, Campbell EA, Hirata Y, Takayama T, Listowsky I. A basis for differentiating among the multiple human Mu-glutathione S-transferases and molecular cloning of brain GSTM5. J Biol Chem. 1993;268(12):8893-8. Epub 1993/04/25.

73. Whitbread AK, Tetlow N, Eyre HJ, Sutherland GR, Board PG. Characterization of the human omega class glutathione transferase genes and associated polymorphisms. Pharmacogenetics. 2003;13:131-44.

74. Washburn MP, Wells WW. Identification of the dehydroascorbic acid reductase and thioltransferase (glutaredoxin) activities of bovine erythrocyte glutathione peroxidase. Biochemical and biophysical research communications. 1999;257:567-71.

75. Hussey AJ, Stockman PK, Beckett GJ, Hayes JD. Variations in the glutathione S-transferase subunits expressed in human livers. Biochimica et Biophysica Acta 1986;874:1-12.

76. Kano T, Sakai M, Muramatsu M. Structure and expression of a human class pi glutathione S-transferase messenger RNA. Cancer Res. 1987;47(21):5626-30. Epub 1987/11/01.

77. Morrow CS, Cowan KH, Goldsmith ME. Structure of the human genomic glutathione S-transferase-pi gene. Gene. 1989;75(1):3-11. Epub 1989/01/30.

78. Moscow JA, Fairchild CR, Madden MJ, Ransom DT, Wieand HS, O'Brien EE, et al. Expression of anionic glutathione-S-transferase and P-glycoprotein genes in human tissues and tumors. Cancer Res. 1989;49(6):1422-8. Epub 1989/03/15.

79. Henderson CJ, Wolf CR, Kitteringham N, Powell H, Otto D, Park BK. Increased resistance to acetaminophen hepatotoxicity in mice lacking glutathione S-transferase Pi. Proc Natl Acad Sci U S A. 2000;97(23):12741-5. Epub 2000/11/01.

80. Menegon A, Board PG, Blackburn AC, Mellick GD, Le Couteur DG. Parkinson's disease, pesticides, and glutathione transferase polymorphisms. Lancet. 1998;352(9137):1344-6. Epub 1998/11/05.

81. Shang W, Liu W-H, Zhao X-H, Sun Q-J, Bi J-Z, Chi Z-F. Expressions of glutathione S-transferase alpha, mu, and pi in brains of medically intractable epileptic patients. BMC Neuroscience. 2008;**9**:67.

82. Webb G, Vaska V, Coggan M, Board P. Chromosomal localization of the gene for the human theta class glutathione transferase (GSTT1). Genomics. 1996;33(1):121-3. Epub 1996/04/01.

83. Pemble S, Schroeder KR, Spencer SR, Meyer DJ, Hallier E, Bolt HM, et al. Human glutathione S-transferase theta (GSTT1): cDNA cloning and the characterization of a genetic polymorphism. The Biochemical journal. 1994;300 (Pt 1):271-6. Epub 1994/05/15.

84. Chen H, Sandler DP, Taylor JA, Shore DL, Liu E, Bloomfield CD, et al. Increased risk for myelodysplastic syndromes in individuals with glutathione transferase theta 1 (GSTT1) gene defect. Lancet. 1996;347(8997):295-7. Epub 1996/02/03.

85. Tan KL, Webb GC, Baker RT, Board PG. Molecular cloning of a cDNA and chromosomal localization of a human theta-class glutathione S-transferase gene (GSTT2) to chromosome 22. Genomics. 1995;25(2):381-7. Epub 1995/01/20.

86. Jang SG, Kim IJ, Kang HC, Park HW, Ahn SA, Yoon HJ, et al. GSTT2 promoter polymorphisms and colorectal cancer risk. BMC cancer. 2007;7:16. Epub 2007/01/26.

87. Board PG, Baker RT, Chelvanayagam G, Jermiin LS. Zeta, a novel class of glutathione transferases in a range of species from plants to humans. Biochemical Journal. 1997;328:929-35.

88. Blackburn AC, Woollatt E, Sutherland GR, Board PG. Characterization and chromosome location of the GSTZ1 encoding the human Zeta class glutathione transferase and maleylacetoacetate isomerase. Cytogenetics and Cell Genetics. 1998;83:109-14.

89. Fernandez-Canon JM, Penalva MA. Characterization of a fungal maleylacetoacetate isomerase gene and identification of its human homologue. The journal of biological chemistry. 1998;273(1):329-37.

90. Fernandez-Canon JM, Penalva MA. Characterization of a fungal maleylacetoacetate isomerase gene and identification of its human homologue. J Biol Chem. 1998;273(1):329-37. Epub 1998/02/07.

91. Information NCfB. Glutathione S-transferase alpha 4 [homo sapiens]. [cited 2012 09/08/2012]; Available from: http://www.ncbi.nlm.nih.gov/gene/2941#general-gene-info.

92. Hubatsch I, Ridderstrom M, Mannervik B. Human glutathione transferase A4-4: an alpha class enzyme with high catalytic efficiency in the conjugation of 4-hydroxynonenal and other genotoxic products of lipid peroxidation. The Biochemical journal. 1998;330 (Pt 1):175-9. Epub 1998/04/16.

93. Curtis JM, Grimsrud PA, Wright WS, Xu X, Foncea RE, Graham DW, et al. Downregulation of adipose glutathione S-transferase A4 leads to increased protein

carbonylation, oxidative stress, and mitochondrial dysfunction. Diabetes. 2010;59(5):1132-42. Epub 2010/02/13.

94. Qian J, Jing J, Jin G, Wang H, Wang Y, Liu H, et al. Association between polymorphisms in the GSTA4 gene and risk of lung cancer: a case-control study in a Southeastern Chinese population. Molecular carcinogenesis. 2009;48(3):253-9. Epub 2008/09/04.

95. Liu C-S, Tsai C-S. Enchanced lipid peroxidation in epileptics with null genotype of glutathione S-transferase M1 and intractable epilepsy. Japanese journal of pharmacology. 2002;90:291-4.

96. Ueda K, Ishitsu T, Seo T, Ueda N, Murata T, Hori M, et al. Glutathione S-transferase M1 null genotype as a risk factor for carbamazepine-induced mild hepatotoxicity. Pharmacogenomics. 2007;8(5):435-42.

97. Fukushima Y, Seo T, Hashimoto N, Higa Y, Ishitsu T, Nakagawa K. Glutathione-S-transferase M1 null genotype and combined GSTM1 and GSTT1 null genotypes are risk factors for increased serum gamma-glutamyltransferase in valproic acid-treated patients. The International Journal of Clinical Chemistry. 2008;389:98-102.

98. Shah RR. Pharmacogenetics in drug regulation: promise, potential and pitfalls. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2005;360(1460):1617-38. Epub 2005/08/13.

99. Tan GM, Wu E, Lam YY, Yan BP. Role of warfarin pharmacogenetic testing in clinical practice. Pharmacogenomics. 2010;11(3):439-48. Epub 2010/04/21.

100. Zhu Y, Shennan M, Reynolds KK, Johnson NA, Herrnberger MR, Valdes R, Jr., et al. Estimation of warfarin maintenance dose based on VKORC1 (-1639 G>A) and CYP2C9 genotypes. Clinical chemistry. 2007;53(7):1199-205. Epub 2007/05/19.

101. You JH, Tsui KK, Wong RS, Cheng G. Potential clinical and economic outcomes of CYP2C9 and VKORC1 genotype-guided dosing in patients starting warfarin therapy. Clin Pharmacol Ther. 2009;86(5):540-7. Epub 2009/07/03.

102. Caraco Y, Blotnick S, Muszkat M. CYP2C9 genotype-guided warfarin prescribing enhances the efficacy and safety of anticoagulation: a prospective randomized controlled study. Clin Pharmacol Ther. 2008;83(3):460-70. Epub 2007/09/14.

103. Yamaguchi-Kabata Y, Shimada MK, Hayakawa Y, Minoshima S, Chakraborty R, Gojobori T, et al. Distribution and effects of nonsense polymorphisms in human genes. PloS one. 2008;3(10):e3393. Epub 2008/10/15.

104. Fu Z, Kim J, Vidrich A, Sturgill TW, Cohn SM. Intestinal cell kinase, a MAP kinase-related kinase, regulates proliferation and G1 cell cycle progression of intestinal epithelial cells. American journal of physiology Gastrointestinal and liver physiology. 2009;297(4):G632-40. Epub 2009/08/22.

105. Berry DJ, Beran RG, Plunkeft MJ, Clarke LA, Hung WT. The absorption of gabapentin following high dose escalation. Seizure : the journal of the British Epilepsy Association. 2003;12(1):28-36. Epub 2002/12/24.

106. Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. The New England journal of medicine. 2010;362(13):1181-91. Epub 2010/03/12.

107. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature. 2008;456(7218):66-72. Epub 2008/11/07.